ajer

THE ALBERTA JOURNAL OF EDUCATIONAL RESEARCH

VOLUME XXXVII NUMBER 3 SEPTEMBER 1991

PUBLISHED BY
THE UNIVERSITY OF ALBERTA • EDMONTON

The Alberta Journal of Educational Research

Published in March, June, September, and December
by the Faculty of Education, University of Alberta

A quarterly journal devoted to the dissemination, criticism, interpretation, and encouragement of all forms of systematic enquiry into education and fields related to or associated with education.

Editor: Thomas O. Maguire

Consulting Editors

Rod Clifton
University of Manitoba

Dianne Common
Lakehead University

Carol Crealock
University of Western Ontario

Janet Donald
McGill University

John Hattie
University of Western Australia

Billie Housego
University of British Columbia

Lilian Katz
University of Illinois

Les McLean
Ontario Institute for Studies in Education

John McLeod
University of Saskatchewan

Linda Phillips
Memorial University of Newfoundland

Lauran Sandals
University of Calgary

Keith Sullivan
Dalhousie University

Phil Winne
Simon Fraser University

Ruth Wright
University of Ottawa

Faculty Advisory Committee

Harvey Zingle (Chair)
Fern Snart
Erwin Miklos
Graham Fishburne

AJER gratefully acknowledges support from the Social Sciences and Humanities Research Council of Canada, from the Alberta Advisory Committee for Educational Studies, and from the Office of Research Services, University of Alberta.

The subscription rate is $32.00 per year for individuals, $40.00 per year for institutions. Add $8.00 for delivery outside Canada. Single copies are $8.00 each. Subscriptions and sales in Canada will be charged 7% GST. Please make cheques payable to *The Alberta Journal of Educational Research*. Back issues are available; rates supplied on request. Claims for undelivered copies must be received within three months of publication.

Address all communications and manuscript submissions to *The Alberta Journal of Educational Research*, Faculty of Education, Publication Services, 4-116 Education North, University of Alberta, Edmonton, Alberta, Canada, T6G 2G5.

Second Class Mail Registration Number 1436

Faculty of Education
University of Alberta

The Alberta Journal of Educational Research

Volume XXXVII, Number 3, September 1991

Special Section

W. TODD ROGERS 193 Introduction to Special Section on Gender Differences in Achievement

SHARON HAGGERTY 195 Gender and School Science: Achievement and Participation in Canada

JEFFREY W. BULCOCK, MORLEY E. WHITT, MONA J. BEEBE 209 Gender Differences, Student Well-Being and High School Achievement

GAALEN ERICKSON, SANDRA FARKAS 225 Prior Experience and Gender Differences in Science Achievement

BIKKAR S. RANDHAWA 241 Gender Differences in Academic Achievement: A Closer Look at Mathematics

Perspectives

JOHN HATTIE 259 The Burt Controversy: An Essay Review of Hearnshaw's and Joynson's Biographies of Sir Cyril Burt

WILLIAM J. HAGUE 277 Kohlberg's Legacy— More than Ideas An Essay Review

ISSN 0002-4805

The Alberta Journal of Educational Research *Vol. XXXVII, No. 3, September 1991, 193-194*

Introduction to Special Section on Gender Differences in Achievement

W. TODD ROGERS
Guest Editor

It is hard to find a more controversial topic in recent behavioral and social science research than that of gender differences in achievement. During the last 30 years or so, a great deal of research has been devoted to explaining these differences. Yet debate persists as to the extent of and reasons for such differences and the validity and significance of the findings and explanations provided.

Much of the research has been conducted in the areas of science and mathematics, likely because of the prominence given to mathematics and science as cornerstones for effective and meaningful economic development and international competitiveness in today's world. Notably, and of concern, is the observation that women are underrepresented, some would argue dramatically, on university science and mathematics faculties and in scientific and technical careers. For many, the reasons for this underrepresentation are traceable to the failure to account for differential performance in school, particularly at the secondary school level, where students begin to select the program of studies they will follow and the science and mathematics courses in which they will enroll, and where there is general agreement that gender differences in achievement begin to emerge.

The four articles included as a special section on gender differences in this issue of *AJER* address the issue of gender differences in science and mathematics during the school years in a Canadian context. Collectively, the articles convey an excitement that research may actually make a difference, that research findings can be used to develop relevant and appropriate procedures for meeting the learning needs of both male and female students in science and mathematics at all levels of schooling. In agreement with research conducted in the United States and elsewhere, there is consistent reference that gender differences in science and mathematics begin to emerge about the onset of adolescence. We must be doing something to young adolescent girls and boys and/or we are failing to take proper account of prior learnings and experiences girls and boys enjoy to socialize them into different patterns of behavior, choice, opportunity, and expectations.

Individually, each article contributes knowledge leading to a greater understanding of gender differences in science and mathematics, and each raises questions that underscore the need for additional research. In the first article, Sharon Haggerty comprehensively surveys the Canadian scene in science and introduces many of the themes that appear in the subsequent articles. Jeff Bulcock, Morley Whitt, and Mona Beebe then address the question "Why are female high school students in Newfoundland outperforming boys in all school subjects except physics?" Rejecting both biological-

ly based theory and culturally based theory, they test a demographic model for accounting for the gender differences in school achievement they observed using causal modeling. In the third article, Gaalen Erickson and Sandra Farkas address the issue of gender differences from a qualitative perspective. They explore the apparent impact of prior experience, in and out of school, on student performance on multiple-choice science achievement items. In particular, they address the question "Why do girls have less success 'applying' concepts to practice than do boys?" Employing the notion of effect size and advanced data analysis procedures, Bikkar Randhawa examines possible reasons for differential effects in mathematics at three different grade (age) levels. To clarify his findings, he simultaneously investigates gender differences in reading using the same analytical procedures.

One disturbing trend seems to be that while researchers are aware of and trying to do something about the issue of gender differences, policy makers and school officials, particularly teachers, may be less convinced. There may be better opportunities and riches for all if the findings of these studies are carefully considered and used by those who work with school students.

A possible disappointment with this special section of *AJER* on gender differences is that the articles, taken as a set, emphasize science and mathematics. Little attention is given to other subject areas such as literature and the fine arts. Three of the four articles were presented in separate sessions of the 1990 Annual Conference of the Canadian Society for the Study of Education/Societé Canadienne pour l'Étude d'Éducation. All the presenters were asked if they wished to have their presentation considered for publication in *AJER* with the understanding that each article would undergo the regular review process. All agreed. Each was asked to nominate one or two other researchers whom he or she knew were working in the area of gender differences. Of the nominees, all were working in the areas of science and mathematics. Does the failure to attract researchers investigating gender differences in other subject areas reflect the absence of such research? While, as mentioned earlier, science and mathematics are key subjects, so are other subjects in the humanities and social science areas. A second question to ask and which arises from the findings presented here and elsewhere is not "Do gender differences occur?" but "Under what circumstances do differences exist and what can be done to provide opportunities for both girls and boys to have their understandings reflected in achievement?" There is a task here for test-makers. We invite you to analyze the findings of the four articles presented here and the literature as we, with our interests in testing, have just done to identify other questions that can be used to direct and clarify the nature and causes of gender differences observed in achievement not only in science and mathematics, but also in the humanities and social sciences.

In closing, we wish to thank the reviewers for their insightful comments and constructive suggestions. The set of articles is stronger as a result of their contributions and concern for gaining greater understanding of gender differences and ways they may be positively addressed in school.

The Alberta Journal of Educational Research *Vol. XXXVII, No. 3, September 1991, 195-208*

SHARON M. HAGGERTY
University of Western Ontario

Gender and School Science: Achievement and Participation in Canada

A review of recent studies of gender differences in science achievement and participation suggests that achievement differences cannot be understood or addressed without considering the effects of both attitude and participation. Canadian findings on achievement and participation support the suggestion that females and males approach school science differently and that those approaches are an important factor in the achievement differences that are observed. Enrollment data from ministries and departments of education reveal marked differences in participation in elective sciences, with the smaller provinces appearing to be more successful at encouraging students to persist in their science studies at the senior high school level.

The issue of gender differences in science and mathematics achievement has probably received more attention than in any other subject. Concern for girls and science has existed internationally since the publication of the results of the first international science assessment (Comber & Keeves, 1973; Kelly, 1978). Arising from this concern, an international organization, Gender and Science and Technology (GASAT), was established in 1981 and continues to meet biennially[1] to work toward achieving educational equity in science, mathematics, and technology. Canadians have participated in GASAT since its inception.

The Science Council of Canada (1984) conducted a major study of science education in Canada. Among its many recommendations were two specifically developed to increase the participation of girls in science. At the provincial level:

> Ministries should outline, in all policy documents relating to science and technology education, concrete steps to be taken by educators to increase the participation of girls in science and technology. (p. 48)

and at the national level:

> The Minister of State for Science and Technology and the Minister responsible for the Status of Women should together sponsor a large-scale public information program directed especially towards parents and designed to raise public awareness of the need for girls and young women to participate in science and technology education. (p. 63)

Attention has also been drawn to gender differences in favor of males in science achievement[2] as measured in various provincial (e.g., Bateson,

---

Sharon Haggerty is an assistant professor of curriculum studies at the University of Western Ontario. Her current research interests are related to perceptions of the nature of science and science teacher education.

Anderson, Dale, McConnell, & Rutherford, 1986) and international assessments (e.g., Connelly, 1987; Connelly, Crocker, & Kass, 1985; Lapointe, Mead, & Phillips, 1989).

While gender differences continue to be reported, debate persists as to their extent and their validity and significance. Not all studies have detected achievement differences; and in studies where differences are found there are a number of views as to the reasons why they occur. Some suggest that gender differences can be accounted for by differential course taking, while others believe the causes are more deeply embedded in factors related to differential socialization of females and males.

As well as achievement and participation in science, attitudes toward science are considered in this article. A recent longitudinal study of influences on attitude toward and achievement in high school science conducted in the United States revealed a strong relationship between achievement and attitude (Simpson & Oliver, 1990). It was found that school variables strongly influenced attitudes, and that attitudes of grade 10 students successfully predicted which students would not pursue future elective science courses in 90% of the cases studied. Hence the view taken in this article is that achievement and attitudes should not be examined in isolation from each other. It is also suggested that merely examining gender differences is little more than an academic exercise. We must also be prepared to develop and implement strategies aimed at ensuring that both females and males in Canada are equally successful in school science. Suggestions as to how that might be addressed are presented.

Achievement in Science

As indicated above, recent science assessments in Canada and elsewhere have found gender differences in favor of males in science achievement, and that these differences first became noticeable at early adolescence, increasing during the high school years. Individual science teachers, however, may dispute their existence, stating there are no differences between the achievement of females and males in their classes (Sadker & Sadker, 1986; Tobin & Garnett, 1987). Moreover, average grade 12 final marks obtained from the British Columbia Ministry of Education reveal little difference between the final marks of male and female grade 12 students in science in that province (Kozlow, personal communication, 1986). Thus there is conflicting evidence as to whether gender differences in science achievement really do exist.

In order to account for this apparent discrepancy, Haggerty (1987) suggested that there may be two distinctly different kinds of achievement measures in science. In the first instance, a test score is being considered. Generally, the test is externally prepared and presented in a multiple choice format. It covers a wide range of content that the students are expected to have encountered in their previous science studies. The students generally do not know what specific content is covered on the test. Further, the context of the items may be quite different than the context in which the particular scientific principles were studied in school, and at times the students do not have the opportunity to study a particular body of content to be tested. Perhaps most importantly the students are aware that the results of such

tests will not influence their marks in any way. The second type of achievement measure is a mark or grade obtained by students for a particular science course in school. Such a mark is a composite of test scores and marks/grades received on various assignments. The tests in this instance are very different in nature. Generally, the content tested has been recently studied in a particular science course. The students study content identified by the teacher. A portion of the mark may be given for work effort in class and for completion of assigned homework. Haggerty suggested these two kinds of achievement measures (i.e., scores on standardized tests and course marks) are actually measuring distinctly different sorts of achievement. To investigate this possibility, she examined data from three British Columbia sources.

In the first analysis, scores on the British Columbia science assessment achievement scales showed superior performance by males for almost all science topics (Bateson et al., 1986; Hobbs, Boldt, Erickson, Quelch, & Sieben, 1979; Taylor, Hunt, Sheppy, & Stronck, 1982). The gender differences were greatest for physical and earth science topics and on questions requiring higher level thinking skills. Moreover, the differences increased with age. In the second analysis, final marks obtained by grade 12 students in British Columbia were examined. In contrast to the assessment results, these final marks are comprised of two equally weighted components: one the mark assigned by the classroom teacher and the other the score on a provincial school leaving examination.[3] It was found that in the sciences the average marks obtained by females ranged from 0.8% less than those obtained by males in chemistry to 2.9% more than males in physics. Thus looking at gender differences on these two kinds of measures (i.e., provincial science assessment achievement scales and blended final marks in biology, chemistry, and physics) yielded different findings.

In the third analysis, Haggerty (1986, 1987) drew on data from a study of a grade 9 science class during a unit on heat and temperature. In the study, scores were obtained on a posttest measure that tested how well students could apply scientific knowledge related to a variety of phenomena[4] they had studied in school. The contexts of the items in this posttest, and in a similar pretest administered prior to instruction, were different than those used by the teacher to teach the unit. And, as in the case of the provincial assessment, the students could not study for this test by learning a particular set of facts and the students knew the results of the test would not be included in their course marks. The pretest revealed no significant gender differences in understanding the heat and temperature concepts tested. Similarly, there were no significant gender differences in the scores obtained by the students on the unit test (set by the teacher and limited to questions that could be answered correctly by a student who had memorized key definitions and other facts), and on combined notebook, homework, and unit test marks assigned by the teacher for the unit as a whole. In contrast, the posttest did reveal significant gender differences. After completing the heat and temperature unit, the boys seemed better able than the girls to apply what they had learned to the novel situations presented in the posttest measure. This finding suggests that males understood the science they learned more effec-

tively than the females, in spite of the marks obtained in their science course. Similar findings have been reported by Bejar and Blew (1981), De Boer (1984), Levin and Fowler (1984), Maccoby and Jacklin (1974), Sadker, Sadker, and Klein (1986), Simpson and Oliver (1990) and Zarega, Haertel, Tsai, and Walberg (1986). Researchers have attempted to identify plausible explanations for why girls receive course marks comparable to those of boys, yet apparently are not able to apply what they learned as well as boys. One suggestion is that better work habits of females provide them with an advantage on school achievement measures (De Boer, 1986; Maccoby & Jacklin, 1974; Walberg, 1969; Zarega et al., 1986). Ridley and Novak (1983) have suggested that females are socialized toward rote learning strategies. Concern for right answers among females is often noted by teachers and was clearly evident in the classroom study conducted by Haggerty (1986). Such findings reinforce the contention that the important question is not whether gender differences exist, but rather, under what circumstances do gender differences exist and, most importantly, what can be done to provide both females and males with opportunities to understand the science they learn and to have that understanding reflected in their achievement?

As noted earlier, there appear to be three plausible explanations as to why males outperform females on standardized tests (such as the British Columbia Science Assessment and the American National Assessment of Educational Progress or NAEP). Eylon and Linn (1989) are among those who support the differential course taking hypothesis, suggesting that because males take more science, they perform better. Erickson and Erickson (1984), however, using data from the British Columbia science assessment, dispute that. Similarly, Mullis and Jenkins (1988), using American NAEP data, found that course enrollment had no effect on the magnitude of the gender difference. In the second international science assessment (International Association for the Evaluation of Educational Achievement, 1988), when students enrolled in their final year of secondary science were tested by subject, there were still significant gender differences in Canada and most other countries (differences for chemistry and physics were greater than for biology).

In their recent review of gender and mathematics and science, Linn and Hyde (1989) conclude that informal as well as formal learning experiences were important. They noted:

> Gender differences in science knowledge and scientific reasoning reflect both formal and informal learning experiences. Overall gender differences in science (a) are larger for science knowledge than for science processes, (b) are declining among high school students, and (c) accompany gender differences in formal and informal learning experiences. (p. 22)

Other British Columbia work by Erickson and Farkas (1987) suggest that males and females tend to draw on different sources of personal knowledge when confronted with questions on a standardized test, and that a lack of relevant prior informal experiences contributes to females' inferior performance. Students were questioned as to what they had been thinking about as they chose their answers to science assessment questions on which sig-

nificant achievement differences between females and males had occurred. Females more than males indicated that they recalled an incident that had occurred in school science classes, whereas males more than females recalled an incident outside of formal school science. Similarly, Linn and Hyde noted that "males compared with females report substantially more informal experience with physical science" (1989, p. 22). Haggerty (1987) has suggested that females may do better on test questions that can be answered by rote learning, whereas the reverse may be true for males. This suggestion is consistent with the findings of both Erickson and Farkas (1987) and Linn and Hyde (1989) and could also account for males' superior performance on higher level assessment questions and for females' superior performance on lower level questions. Similarly, school science course marks are more likely to reflect good work habits and to be based on tests that ask predominantly lower level questions, thereby contributing to the relatively higher achievement of females on these measures.

One additional explanation has been proposed to account for the differences on British Columbia assessment scales. Bateson and Parsons-Chatman (1989) suggest that the differences may be due to gender bias in the content of the items. Erickson and Farkas's (1987) findings do not support that explanation. At best, the content of the items appears to be insufficient to account for the explanations given by students in Erickson and Farkas' study.

The view that any one factor such as differential course taking, prior informal experiences, or content of items is responsible for gender differences in achievement is not supported by the studies reviewed above. There is now considerable evidence that achievement is influenced by a complex of factors.

Attitudes Toward Science

In a meta-analytic study of gender and science, Steinkamp and Maehr (1984) concluded that although in general females were more motivated to achieve than males, the reverse was true at the junior high level in the United States. There, adolescent males' motivation to succeed exceeded that of females and the gap between the two had been increasing during the preceding decade. Since then, Simpson and Oliver (1990) have conducted a longitudinal study with adolescents in the United States. Using extensively revised existing instruments, as well as developing additional instruments for their work, they finally produced and utilized a 60-item, 14-subscale attitude survey. In their study, females scored consistently higher on motivation to achieve in science (among other subscales), whereas males had significantly more positive attitudes toward science. Simpson and Oliver noted, as have others (e.g., Duncan & Haggerty, 1985; Yager & Yager, 1985), a general decline in attitudes toward science during the adolescent years. In a Saskatchewan study of 43 grade 7 to 9 science classes, Wideen (1974) found that interest in science was greater for males than females, and that for students who indicated that a subject other than science was their favorite subject, there was a decline in science interest from grade 7 to grade 9. The decline did not occur among students who chose science as their favorite subject in school.

Similarly, Kelly (1986) included gender as a factor in a longitudinal study of science attitudes. She found that from ages 11 to 13 children's attitudes toward science declined significantly and noted that "their personal liking for science showed a particularly marked drop, especially among girls" (p. 402). These declines in adolescents' attitudes toward science should be of concern to those involved in science education during the middle school/junior high school years.

Linn and Hyde (1989) discussed three aspects of attitude (or what they call *psychosocial*) differences: aggression (e.g., males' greater participation in class activities and discussions), confidence (e.g., males overestimate their abilities, females are realistic), and interest. They noted that gender differences were heterogeneous and interacted both with the particular situation and with cultural factors. Linn and Hyde concluded that the only interest factor that seemed to influence persistence in science and mathematics was perceived usefulness, and suggested that this may be related to the frequently reported concern that the science taught in schools is not seen as addressing relevant topics or problems. As most provinces have expressly developed curricula with relevance to students (particularly at the middle school/junior high school level), this is a concern. It appears that although curriculum developers believe the curriculum addresses topics that are socially relevant to adolescents, that relevance is not apparent to the students. If Linn and Hyde are correct that science participation is related to perceived relevance, then perhaps we might hope to increase participation if we can make that relevance apparent to the students. Similarly, we might look to perceived relevance as we ponder explanations for low participation in elective science courses.

Simpson and Oliver (1990) have come to somewhat different conclusions. Data were initially collected on 4,500 grade 10 students. To determine the extent to which further participation in science could be predicted, additional data were collected five years later from these students. Student attitudes toward science and success in science in grade 10 were a strong predictor of subsequent science achievement in high school, and science self concept at grade 10 was a significant predictor of future course participation. In particular, students with low science attitudes did not take elective science courses. The authors concluded, "A major finding of this study, therefore, is that attitudes toward science play a key role in influencing the amount of exposure to science a student experiences" (p. 13). The longitudinal nature of this study appears to have allowed more precise identification of the relationship between attitude and both achievement and participation in elective sciences. Simpson and Oliver also found, contrary to their expectations, that school variables were more strongly related to attitude toward science than were family variables. This is encouraging for educators who are sometimes discouraged from attempting to address student attitudes due to the perception that the home is the major influence on student attitudes. Simpson and Oliver's finding suggests that it is important to consider the school environment when addressing attitude issues.

Participation in Science

Two aspects of participation are considered—choosing to enroll in elective science courses and student participation in discussions and activities in the science classroom.

Enrollments in Elective Science

Studies reviewed in the previous section suggest that attitude is an important factor in a student's choice to enroll in science courses. Ridley and Novak (1983) have suggested that the socialization of females toward rote learning strategies may be related to low female participation in advanced science and mathematics courses. In order to examine the pattern of student enrollments in science, representatives of each of the 10 provincial departments/ministries of education were asked to provide student enrollment data for grade 10 to 12 science courses. In the first analyses of these data, male and female enrollment in grade 11 sciences were compared with the corresponding populations in each province. Generally, grade 11 marks the point at which enrollment in a science course becomes optional rather than compulsory, that is, science becomes an elective rather than a required course. To obtain an estimate of the total number of females and males who might potentially enroll in a grade 11 science, the population of 15- to 19-year-olds (Statistics Canada, 1988) was divided by five. There is greater participation by both males and females in some provinces than in others. For example, in Saskatchewan the percentages of females enrolled in biology, chemistry, and physics were respectively, 76%, 61%, and 47%. The corresponding data for Ontario were 43%, 32%, and 20%. This result illustrates a second point: higher enrollments in grade 11 science tend to be inversely related to total grade 11 population size across the six provinces for which there were sufficient data (see Table 1). Thirdly, examination of the corresponding

Table 1
Estimated Participation of Females and Males in Grade 11 Biology, Chemistry, and Physics by Province
1987/1988

Province	Female Pop'n X1000 (15-19)	Females			Male Pop'n X1000 (15-19)	Males		
		Biol.	Chem.	Phys.		Biol.	Chem.	Phys.
B.C.	101.8	45%	34%	19%	106.9	40%	34%	34%
Alberta	89.3	67	58	24	92.9	54	55	38
Saskatchewan	38.3	76	60	47	40.3	64	55	50
Ontario	336.4	43	31	20	352.7	29	34	28
N.B.	29.4	34	40	28	31.0	30	36	33
P.E.I.	5.2	73	60	30	5.6	49	46	39

Note: Data not included for some provinces because of insufficiency (Manitoba and Newfoundland/Labrador) or because not all sciences are offered at the grade 11 level (Nova Scotia and Quebec).

course percentages for females and males reveals that across all provinces a greater percentage of females than males enroll in biology; in contrast, the percentage of females in physics is lower. The pattern of differences for chemistry is less clear.

To further examine the relative participation of females and males enrolled in each science course offered in grades 10, 11, and 12 in each province, the proportion of males and females enrolled in each subject is shown in Table 2. A number of observations can be made. First, in agreement with the data provided in Table 1, more males than females study biology, and, except among francophones in Ontario and Quebec, more males than females studied physics. For chemistry, the proportions of males and females were more nearly equal. There were more males in chemistry in British Columbia and Newfoundland and among central Canadian anglophones, but more females among the francophones and in the maritime provinces. The greater participation by females in francophone classes is interesting and probably warrants further examination.

Although it is impossible to explain these course enrollment variations with certainty, some patterns have been identified and some speculations can be made. The smaller, more rural regions of Canada (i.e., the prairies and Atlantic provinces) tend to have higher science participation. This is particularly so for Saskatchewan and for females in Prince Edward Island. New Brunswick, however, is an exception. New Brunswick differs from the others in that it offers a greater variety of academic and nonacademic courses. In New Brunswick, students who will be continuing to postsecondary education take different courses than those who will not. New Brunswick also has a relatively higher proportion of females in physics than any other province except Saskatchewan (not including francophone classes) and has made an effort to attain gender equity in education (Department of Education, 1987). Physics participation by females is particularly dismal in the final year of schooling in British Columbia, Ontario, and Alberta. These three provinces are more industrialized and economically advantaged than the others.

The differences between Saskatchewan and Alberta are of interest. In both of these provinces approximately half of those enrolled in chemistry are female. In Alberta, however, the typical gender bias (more females take biology, more males take physics) is more pronounced in biology and physics. Historically, these provinces have much in common. There are differences in the dominant economic base and the dominant political philosophy of these neighboring provinces. In Alberta the oil industry is particularly significant in the provincial economy, whereas farming dominates in Saskatchewan.

Overall, among anglophones the gender bias is most pronounced in the three provinces that not only have the larger populations, but also are more industrialized than the others (i.e., Alberta, British Columbia, and Ontario)—the question is, is this merely a coincidence, or is there any possible significance to such trends?

Differences in course participation across Canada illustrate the complexities involved in this issue. Sociocultural influences are undoubtedly involved. However, it behooves those of us in the educational community to focus on school and curriculum variables in the hope that we can have some

Table 2
Percentage of Females in Canadian Secondary Science Classes, 1987/88

Province	Science	Biology			Chemistry			Physics		
Grade:	10[a]	11	12	13	11	12	13	11	12	13
B.C.	50	57	61		49	43		35	21	
Alberta	53/50/49[b]	55	59		50	51		38	33	
Saskatchewan	50	53	55[c]		51	52		47	44	
Manitoba[d]		(57)			(49)			(36)		
Ontario Fr.[e]	56	63	50	63	53	58	47	53	52	41
Eng.	51	59	61	59	48	48	44	43	39	32
Quebec Fr.	53/50[f]				53					
Eng.	49/38				45		47			
N.S.	50/46[g]		58		52	52		45	36	
P.E.I.	54	60	61		55	56		42	43	
N.B: Fr.		68[h]	63		53	57		48	47	
Eng.		53	58		51	51		45	44	
Nfld/Lab.[i]		60			43			27		

[a]In most provinces general science is compulsory in grade 10.

[b]In Alberta in 1987/88 each of biology, chemistry, and physics was offered in grade 10. Each student was required to take at least *one* of the three courses. In 1990/1991 general science will replace the three separate courses.

[c]Francophone biology 12 classes were 57% female.

[d]Data for Manitoba were estimated from a bar graph; grade was not indicated.

[e]Ontario, Quebec, and New Brunswick maintain separate records for francophone and anglophone classes. Ontario is the only province that provides a 13th year of schooling.

[f]In Quebec there is no grade 12. Chemistry and physics are offered at years 10 and 11, respectively. Biology is not offered after year 9.

[g]In Nova Scotia biology and physical science are offered in grade 10.

[h]The first year of biology in New Brunswick is a grade 10 course.

[i]Newfoundland/Labrador did not respond to the request for data. Figures given are taken from Scott (1982) and are for the final year of secondary school which at that time was grade 11. Since then grade 12 has been introduced as well.

impact on how students experience school science. The following section examines how the school science experience differs for females and males.

Participation in Science Classes

Differences in the way females and males interact and respond in the classroom have been well documented in a variety of subject areas and at all age levels (Becker, 1981; Bossert, 1981; Haggerty, 1987; Linn & Hyde, 1989; Sadker & Sadker, 1986; Tobin & Gallagher, 1987; Wienekamp, Jansen, Fickenfrerichs, & Peper, 1987; Whyte, 1984). For example, females appear less likely to engage in high risk situations (Lolley, Davies, & Scott-Hodgetts, 1986). And, although females are more likely to respond "I don't know," they are less likely to give wrong answers than males (Linn, De Benedictis, Delucchi, Harris, & Stage, 1987).

In Haggerty's (1986) classroom study of a grade 9 science class studying heat and temperature, the frequency of different forms of teacher-student verbal interactions was recorded during two weeks of class discussion (Table 3). Although only 39% of the students in the class were male (9 of 23), males were called upon to answer 58% of the questions and males initiated 67% of the student-initiated dialogue (see Table 3). When students did initiate dialogue, females were more likely to provide information, whereas males were more likely to ask a question. Although females comprised 61% of the class, they responded to only 42% of the teacher's questions. However, 75% of their responses were accepted as correct by the teacher. The males, who responded to 58% of the questions asked by the teacher, provided correct responses 64% of the time. When the teacher responded to students' responses, she was more likely to encourage further responses or to provide explanations when responding to a female student, and less likely to redirect a question or ignore a response from a female student. Taken together, these two forms of response (student, teacher) suggest that in spite of the teacher's positive and encouraging responses to female students, females were still much less responsive than the males. These findings are consistent with those reported in other studies in the United States and Australia (Sadker & Sadker, 1986; Tobin & Gallagher, 1987).

Kahle, Matyas, and Cho (1985), Tobin and Garnett (1987), and Zimmerer and Bennett (1987) all found that males reported different perceptions of their activities in school science than did females in the same classes. Males reported greater time spent in activities than did females in the same classes. Haggerty (1987) suggested these differing perceptions may be related to the ways in which males and females go about doing science. Males appear to be more involved in manipulative laboratory activities and females to be more concerned with finding the right answers.

The issue of male dominance in classrooms is a difficult one. In classes taught by teachers who claimed to actively encourage participation of female students, males still dominated the activities and discussions, a situation that was not recognized by the teachers concerned (Duncan & Haggerty, 1985; Sadker & Sadker, 1986; Tobin & Garnett, 1987). Observers noted that even when a teacher deliberately called on students proportionally, males were more likely to attempt a response when they were unsure about a question. This provided an advantage for the males, as even an incorrect answer provides the teacher with a basis for continuing to probe and draw out a more acceptable response. Alternatively, a response of "I don't know" gives the teacher nothing to build on and she or he typically passes on to another student. If we wish to overcome this problem, teachers will need to develop strategies to draw out and encourage all students, both female and male, to make an attempt to answer questions. Teachers must convince students (and be convinced themselves) that giving the correct answer should not be the major goal in science classrooms. Rather, the process of developing and evaluating alternative answers is the essence of doing science. As Tobin and Garnett (1987) have noted, the first step is for teachers to recognize the tendency for students to want to identify a single correct answer to any question as problematic.

Table 3
Student-Teacher Dialogue During Class Discussion

	Male		Female		Total
	N	%	N	%	N
Number of Students	9	39.1	14	60.9	23
Students Called on by Teacher	43	58.1	31	41.9	74
Student Responses to Questions					
Correct Answer	100	63.7	86	75.4	186
Incorrect Answer	14	8.9	13	11.4	27
Partially Correct	13	8.3	10	8.8	23
Alternative Belief	21	13.4	6	5.3	27
No Response	6	3.8	2	1.8	8
Total	157	100.0	114	100.0	271
Student Initiated Dialogue					
Question	37	55.2	12	36.4	49
Information	30	44.8	30	60.6	50
Total	67	100.0	32	100.0	99
Teacher Responses to Students					
Encourage, Explore	100	33.9	71	38.2	171
Acknowledge Answer	65	22.0	44	23.7	109
Wrong Answer	11	3.7	2	1.1	13
Provide Information	38	12.9	19	10.2	57
Explanation	28	9.5	27	14.5	55
Demonstration	2	0.7	0		2
Redirect Question	7	2.4	1	0.5	8
Repeat	24	8.1	19	10.2	43
Ignore or Dismiss	16	5.4	2	1.1	18
Managerial	4	1.4	1	0.5	5
Total	295	100.0	186	100.0	481

Conclusions

This article reviews recent studies of gender differences in science achievement, attitude, and participation. Achievement differences are often considered to be related to attitudes toward science and to different participation rates in science. The studies reviewed suggest that these three factors are interrelated and that each contributes to the other.

At the present time, although participation varies across Canada, generally fewer than half of our students choose to study elective science courses. Differences in participation of francophone and anglophone females were unexpected. Some sociocultural factors that might possibly influence participation have been speculated on, but these would require considerably more investigation before any claims could be made.

Studies completed in Canada and elsewhere are reviewed in this article and suggest that for females more than for males success in school science is attained by learning right answers. It is suggested that the result of this approach to learning is that girls more than boys frequently do not understand the science they learn. This approach to learning may contribute to females' lower level of achievement on standardized tests. In turn, failure to understand the science that is being learned leads females to feel that science, particularly physics, is not for them. Consequently, they are less likely to participate in elective physical science courses. If females' disadvantage with respect to both participation and achievement is to be overcome, it will be necessary for middle school/junior high school science to be taught in a more meaningful and relevant way, involving girls as well as boys as active participants in science classes. Oliver and Simpson (1988) stress the importance of "structuring the science classroom learning environment to improve the student's belief that he or she can succeed in science" (p. 154). Parker and Rennie (1986) found that teacher in-service aimed at reducing gender stereotyping of science had positive effects on 10-year-olds in Australia. Kelly (1986) reported that interventions aimed at improving attitudes of 11- to 13-year-olds in several British schools also had positive effects. Duncan and Haggerty (1985) found that in one of two Canadian high schools they studied, the science department head/chemistry teacher successfully recruited large numbers of female students into grade 11 chemistry. Informal anecdotal information from teachers also suggests that individual, enthusiastic teachers can make a difference. Studies currently underway in British Columbia are aimed at identifying characteristics of schools that do have above average participation of females in science. Hopefully, this work will assist teachers and schools to identify ways of increasing participation in science.

Overall, the findings from these studies are encouraging, and teachers should feel empowered by such research. Teachers can have an impact on achievement and participation of their students in school science.

Acknowledgment

The author would like to thank Phil Nagy, the anonymous reviewers, and particularly Todd Rogers for their many helpful suggestions.

Notes

1. The 1993 GASAT conference will be held in Canada at the University of Waterloo. Only once before has GASAT met in North America (at the University of Michigan in 1987).
2. Although this article specifically deals with science, many of the situations and views described apply equally to mathematics.
3. At the time of this study, grade 12 final marks in academic courses were determined by a combination of a teacher-assigned mark and the score obtained on the provincial examination. The two scores were equally weighted.
4. For example, students were told that if there were a large ice cube and a small ice cube in a beaker of water, the large ice cube would take longer to melt than the smaller one. They were then asked if that meant it took more heat to melt the larger cube. On both the pretest and the posttest many students replied no, because ice always melts at 0 degrees. Such students clearly did not distinguish heat and temperature appropriately, although that was a major goal of the unit, and all students were able to define the two terms correctly on the unit test.

References

Bateson, D., Anderson, J., Dale, T. McConnell, V., & Rutherford, C. (1986). *British Columbia science assessment: General report*. Victoria, BC: Ministry of Education.

Bateson, D.J., & Parsons-Chatman, S. (1989). Sex-related differences in science achievement. *International Journal of Science Education, 11*, 371-385.

Becker, J.R. (1981). Differential treatment of females and males in mathematics classes. *Journal for Research in Mathematics Education, 12*, 40-53.

Bejar, I.I., & Blew, E.O. (1981). Grade inflation and the validity of the Scholastic Aptitude Test. *American Educational Research Journal, 18*, 143-156.

Bossert, S.T. (1981). Understanding sex differences in children's classroom experiences. *Elementary School Journal, 81*, 255-266.

Comber, L.C., & Keeves, J.P. (1973). *Science education in nineteen countries*. Stockholm: Almquist and Wiksell.

Connelly, F.M. (1987). *Ontario science education report card: Canadian national comparisons*. Toronto, ON: Queen's Printer.

Connelly, F.M., Crocker, R.K., & Kass, H. (1985). *Science education in Canada: Policies, practices and perceptions*. Toronto, ON: Ontario Institute for Studies in Education.

De Boer, G.E. (1984). A study of gender effects in the science and math course-taking behavior of a group of students who graduated from college in the late 1970's. *Journal of Research in Science Teaching, 21*, 95-103.

De Boer, G.E. (1986, April). *Perceived science ability and effort among males and females: Relationships to persistence in undergraduate science studies*. Paper presented to the Annual Meeting of the American Educational Research Association, San Francisco.

Department of Education. (1987). *Educational opportunities for girls and boys: A commitment to sex-equity*. Fredericton, NB: Author.

Duncan, M.J., & Haggerty, S.M. (1985). *Action plan for girls in science in British Columbia*. Report prepared for the Department of the Secretary of State, Canada, Women's Programs.

Erickson, G.L., & Erickson, L.J. (1984). Females and science achievement: Evidence, explanations, and implications. *Science Education, 68*, 63-89.

Erickson, G., & Farkas, S. (1987). Prior experience: A factor which may contribute to male dominance in science. In J.S. Daniels & J.B. Kahle (Eds.), *Contributions to the Fourth GASAT Conference* (vol. 1, pp. 49-56), Ann Arbor, MI: University of Michigan.

Eylon, B., & Linn, M.L. (1989). Learning and instruction: An examination of four research perspectives in science education. *Review of Educational Research, 58*, 251-301.

Haggerty, S.M. (1986). *Learning about heat and temperature: A study of a grade 9 science class*. Unpublished doctoral dissertation, University of British Columbia.

Haggerty, S.M. (1987). Gender and science achievement: A case study. *International Journal of Science Education, 9*, 271-279.

Hobbs, E.D., Boldt, W., Erickson, G., Quelch, T., & Sieben, G. (1979). *British Columbia science assessment: General report*. Victoria, BC: Ministry of Education.

International Association for the Evaluation of Educational Achievement. (1988). *Science achievement in seventeen countries*. Oxford: Pergamon.

Kahle, J.B., Matyas, M.L., & Cho, H. (1985). An assessment of the impact of science experiences on the career choices of male and female biology students. *Journal of Research in Science Teaching, 22*, 385-394.

Kelly, A. (1978). *Girls and science*. Stockholm: Almquist and Wiksell.

Kelly, A. (1986). The development of girls' and boys' attitudes to science: A longitudinal study. *European Journal of Science Education, 8*, 399-412.

Lapointe, A.E., Mead, N.A., & Phillips, G.W. (1989). *A world of differences: An international assessment of mathematics and science*. Princeton, NJ: Educational Testing Service.

Levin, J., & Fowler, H.S. (1984). Sex, grade, and course differences in attitudes that are related to cognitive performance in secondary science. *Journal of Research in Science Teaching, 21*, 151-166.

Linn, M.C., De Benedictis, T., Delucchi, K., Harris, A., & Stage, E. (1987). Gender differences in National Assessment of Education Progress items: What does "I don't know" really mean? *Journal of Research in Science Teaching, 24*, 267-278.

Linn, M.C., & Hyde, J.S. (1989). Gender, mathematics, and science. *Educational Researcher, 18*(8), 17-19, 22-27.

Lolley, M., Davies, S., & Scott-Hodgetts, R. (1986). Teachers as researchers. *Mathematics Teaching, 118*, 46-47.

Maccoby, E.E., & Jacklin, C.N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.

Mullis, I.V.S., & Jenkins, L.B. (1988). *The science report card, elements of risk and recovery: Trends and achievement based on the 1986 national assessment* (Report No. 17-S-01). Princeton, NJ: Educational Testing Service.

Oliver, J.S., & Simpson, R.D. (1988). Influences of attitude toward science, achievement motivation, and science self concept on achievement in science: A longitudinal study. *Science Education, 72*, 143-155.

Parker, L.H., & Rennie, L.J. (1986). Sex stereotyped attitudes about science: Can they be changed? *European Journal of Science Education, 8*, 173-183.

Ridley, D.R., & Novak, J.D. (1983). Sex-related differences in high school science and mathematics enrolments: Do they give males a critical headstart toward science- and math-related careers? *Alberta Journal of Education, 29*, 308-318.

Sadker, D., & Sadker, M. (1986). Sexism in the classroom from grade school to graduate school. *Phi Delta Kappan, 56*, 512-515.

Sadker, D., Sadker, M., & Klein, S.S. (1986). Abolishing misconceptions about sex equity in education. *Theory into Practice, 25*, 219-226.

Science Council of Canada. (1984). *Science for every student: Educating Canadians for tomorrow's world*. Ottawa: Author.

Scott, J. (1982). Is there a problem? In J. Ferguson (Ed.), *Who turns the wheel?* (pp. 21-44). Ottawa: Science Council of Canada.

Simpson, R.D., & Oliver, J.S. (1990). A summary of major influences on attitude toward and achievement in science among adolescent students. *Science Education, 74*, 1-18.

Statistics Canada. (1988). *Canada yearbook: 1988*. Ottawa: Supply and Services Canada.

Steinkamp, M.W., & Maehr, M.L. (1984). Gender differences in educational motivations toward achievement in school science: A quantitative synthesis. *American Educational Research Journal, 21*, 39-59.

Taylor, H., Hunt, R., Sheppy, J., & Stronck, D. (1982). *British Columbia science assessment: General report*. Victoria, BC: Ministry of Education.

Tobin, K., & Gallagher, J.J. (1987). The role of target students in the science classroom. *Journal of Research in Science Teaching, 24*, 61-75.

Tobin, K., & Garnett, P. (1987). Gender related differences in science activities. *Science Education, 71*, 91-103.

Walberg, H.J. (1969). Physics, femininity, and creativity. *Developmental Psychology, 1*, 47-54.

Whyte, J. (1984). Observing sex stereotypes and interactions in the school lab and workshop. *Educational Review, 36*, 75-86.

Wideen, M.F. (1974). Division III science: An in depth evaluative study. *Saskatchewan Journal of Educational Research and Development, 5*(1), 15-20.

Wienekamp, H., Jansen, W., Fickenfrerichs, H., & Peper, P. (1987). Does unconscious behaviour of teachers cause chemistry lessons to be unpopular with girls? *International Journal of Science Education, 9*, 281-286.

Yager, R.E., & Yager, S.O. (1985). Changes in perceptions of science for third, seventh, and eleventh grade students. *Journal of Research in Science Teaching, 22*, 347-358.

Zarega, M.R., Haertel, G.D., Tsai, S., & Walberg, H.J. (1986). Late adolescent sex differences in science learning. *Science Education, 70*, 447-460.

Zimmerer, L.K., & Bennett, S.M. (1987, April). *Gender differences on the California statewide assessment of attitudes and achievement in science*. Paper presented to the Annual Meeting of the American Educational Research Association, Washington, DC.

The Alberta Journal of Educational Research *Vol. XXXVII, No. 3, September 1991, 209-224*

JEFFREY W. BULCOCK
Memorial University of Newfoundland

MORLEY E. WHITT
E. Vaders Pentecostal College, St. John's

and

MONA J. BEEBE
Memorial University of Newfoundland

Gender Differences, Student Well-Being and High School Achievement

The study focuses on the question: Why are female high school students in Newfoundland outperforming boys in all school subjects except physics? Support for two theories—a biologically based theory and a culturally based theory—is reviewed and the arguments in favor of both are rejected. Instead, a demographic or cohort size hypothesis was entertained that generated three hypotheses: namely, that under circumstances when relative incomes are low (a) gender effects on the quality of school life will favor girls, (b) gender effects on student well-being will be mediated by the quality of school life and will favor girls, and (c) gender effects on school achievement will be mediated by the quality of school life and will, again, favor girls. While support for the demographic model of gender differences in school achievement was equivocal, the results indicated that student perceptions of their school lives influenced both their well-being and their achievement in literacy and numeracy.

Introduction

This study addressed the question: Why are Newfoundland girls outperforming Newfoundland boys in virtually all school subjects? For the past four years these gender differences have occurred at the grade 4, 6, 8, and 12 levels in the Newfoundland school system (Galway & Banfield, 1991, pp. 22-24; Department of Education, 1988, 1989, 1990, 1991).

For nearly a century there has been continuing research into the nature and origins of gender differences in achievement and ability. As far back as 1873, Herbert Spencer, in an article entitled "Psychology of the Sexes," argued in support of the theories of Charles Darwin that the intellectual attributes of women developed differently in the course of evolution (cited in Keeves, 1985). According to Sadker, Sadker, and Klein (1991), this turn of

---

Jeffrey Bulcock is a professor of education specializing in the social demography of education and research methodology.

Morley Whitt is a teacher in mathematics and physics with a research interest in issues in educational administration.

Mona Beebe is an associate professor of reading and language arts education. Her research interests include reading comprehension in the early grades, adult literacy, and the evaluation of reading instruction methodologies.

the century view that biology determined ability and personality implied that the range of roles open to women was more limited than that open to men (p. 307). In other words, although personality consists of the internalization of roles, roles themselves were subject to organic constraints that happened to favor men. In 1906, E.L. Thorndike rejected the view that gender differences in ability could be inherent on the grounds that the observed differences were too small to be of practical significance (cited in Keeves, 1985). Hollingworth, a student and colleague of Thorndike's, contended that the small differences observed were due to social influences and not to biological causes, and that the true intellectual potential of women would only be revealed when women received a similar education and had the right to choose equivalent careers (Hollingworth & Montague, 1914-1915). These views were congruent with those of the statistician Pearson (1897) who found few significant differences between men and women on a range of variables.

During the 1920s and 1930s research into gender differences was primarily concerned with the study of differences in attitudes and emotional needs. An important shift in research emphasis began in the 1950s. The assumption that roles were biologically given gave way to the thesis that the concept of role was integrally related to the patterns of behavior associated with a given status and position. Status and position in turn consist of expectations or systems of rights and duties (Sarbin, 1954, pp. 224-225). Thus to change roles was to change social expectations or the constellation of rights and duties associated with societal positions.

Maccoby's (1966) synthesis of research on gender differences in the United States, along with those of Tyler (1956) and Anastasi (1958), found that boys generally showed stronger numerical and spatial abilities and performed better on tests of mathematical reasoning than girls, but girls usually did better in verbal and linguistic studies. Later, Tyler (1969) specifically pointed out that when batteries of achievement tests were used to assess achievement, girls exceeded boys in performance in language studies and boys tended to perform better in mathematics and science. She emphasized, however, that these differences were not large, especially when compared with the differences within each gender group.

These findings from the 1960s were reinforced by Maccoby and Jacklin (1974) a decade later. It appeared, paradoxically, that while gender differences in the conative and affective domains were myths, gender differences in the cognitive domain were not. Elaborate explanations were formulated and tested to account for differences in intellectual performances. Especially prominent were hormonal theories (Geary, 1989; Resnick, Berenbaum, Gottersman, & Bouchard, 1986) and theories of hemispherical dominance (Dimond & Beaumont, 1974; Seward & Seward, 1980; Springer & Deutsch, 1981; Wheatley, Mitchell, Frankland, & Kraft, 1978).

The biological basis of sex role differences in cognition still remains a viable explanation. Recent meta-analytic reviews of the gender differences literature in mathematics by Hyde (1981), Feingold (1988), and Marsh (1989), however, show that the differences are small and declining. Ethington (1990), in an analysis of grades 7 and 8 mathematics data from eight countries, concluded that there were no gender effects and, further,

that what effects were discernible were in favor of girls. The same trend seems to hold in the case of science. In their meta-analytic review of the literature on gender differences in science, Linn and Hyde (1989) concluded that the gender gap is disappearing. They argue that gender differences in science reflect parallel differences in course enrollment and training, and that these differences are consistent with changing opportunity, changing social roles, and changing demands in the work place. Hollingworth's turn of the century prediction is being verified. In 1981, gender differences in math and science were the focus of a workshop entitled "Who Turns the Wheel?" dealing with the science education of women in Canada. The report on the workshop (Ferguson, 1982) indicates that studies in other countries have suggested that the separation of girls from science and mathematics starts early in life and that the gap becomes more obvious at each succeeding stage of the education process. As a result, few women graduate from university science and engineering faculties or from polytechnical institutes because they are not entering them. It may be that women do not enter these faculties because they are receiving inadequate science or mathematics training in high school. And this in turn may be a consequence of inadequate preparation and encouragement in the science areas at the primary and elementary school levels.

Physics courses in high school seem to have the least number of girls enrolled of any science/math course. Chemistry is studied more often than physics, and in biology courses there are often more girls than boys. Biology tends to be perceived by students as the least difficult science, and so it is taken by many girls merely to satisfy the science requirement for graduation. In advanced mathematics and pure mathematics courses often less than one third of the class are girls. In addition, by the final year of high school as many as one quarter of the female population in some provinces are no longer studying mathematics and less than half study physics or chemistry at that level. In contrast, 96% of the boys take mathematics in their final year of high school (Ferguson, 1982). Nevertheless, Bulcock and Mendoza (1988) found that when girls enroll in the more difficult science and math courses, they do as well as and sometimes better than boys. The interpretation of these data is complicated by the fact that the small percentage of girls who enroll in physics and chemistry may also represent the most academically able group. The larger percentage of boys is likely to include a significant proportion of individuals of more than average ability. Nonetheless, the achievement figures suggest that girls are capable of doing well in physics, chemistry, biology, and mathematics (Linn & Hyde, 1989).

Most of the data available on gender differences in achievement for Newfoundland are reported in the Department of Education statistical reports for the years 1988 through 1991. At grades 4 (Department of Education, 1991, Table 26), 6 (Department of Education, 1989, Figure 10), and 8 (Department of Education, 1990, Figure 14), girls outperformed boys in virtually all skill areas. These include reading comprehension, mathematics, and overall composite scores. The exception seems to be vocabulary at grade 8.

These findings prevail in grade 12 subjects with the exception of physics (Department of Education, 1988, Table 20; 1989, Table 23; 1990, Table 36). Although many of these differences are modest and therefore unlikely to be statistically significant, the differences in the verbal subjects such as language, literature, the New Testament, and French are substantial. It is necessary to point out, however, that the differences may be underestimated because there are, again with the exception of physics, more female students writing public examinations in these subjects than male students. For example, in the case of Language 3101, a compulsory subject, 4,121 female students wrote the final examination in 1989 compared with 3,517 male students (Department of Education, 1990, Table 36). If the top 3,517 female students were compared with the same number of male students, average achievement score differences would undoubtedly be higher than those found in the Department of Education report.

This section of the article presents two theories that purport to explain gender differences in school achievement. The first theory holds that cognitive behaviors are constrained by innate psychological or organic dispositions, whereas the second accounts for gender differences on the basis of cultural and social influences. However, the Newfoundland data on school achievement are at odds with traditional findings regarding gender differences and, because girls outperform boys in just about every subject in grades 4, 6, 8, and 12, the data do not support the traditional neo-Darwinist explanations. Further, there appears to be no obvious cultural or sociological explanation for these gender differences.

In the remainder of this article, the gender gap reversal in school achievement summarized above is reexamined from a demographic perspective followed by the testing of selected hypotheses derived from this perspective. It is emphasized from the outset that the use of a demographic perspective is exploratory and that the models formulated and tested are conjectural rather than confirmatory. Other influences such as changing societal expectations, expanding educational opportunities for women, and the perceived need for two incomes in families are certain to have contributed to the reversed gender differences observed in Newfoundland.

Demographic Perspective

The demography of human motivation, life style, income levels, and other social variables was developed by Easterlin (1980, 1987a, 1987b) over the past three decades and is referred to as the cohort size hypothesis. This theory holds that the fortunes of a birth cohort (those born in a given year) tend to vary inversely with its relative size; that is, if the birth rate is high adverse social effects will follow as a consequence of demographic crowding. The three institutions most affected by crowding are the family, the school, and the labor market.

During the depression years of the 1930s, birth rates were low. When this generation entered the labor market in the period 1950-1960, the supply of younger workers relative to older workers was low. This ratio of younger workers to older workers governs relative income such that when the ratio is low, relative incomes will be high. When the relative incomes of young men

(compared to the past incomes of their fathers) rise, they feel freer to marry and have children early in adulthood. If relative incomes fall the attendant economic distress will be reflected in declining marriage and fertility rates.

The evidence shows that the impact of declining relative income will be greater for those with high levels of education—for example, those with college degrees—hence there is less incentive for young men to invest in intellectual competencies when relative incomes are low. In contrast, when relative incomes are high, investments in schooling are profitable for young men. But high relative incomes promote high marriage rates and, because young men marry younger women, some women may forgo college for marriage, thereby retarding the growth of female college enrollments to some degree. The dynamics of these relationships between cohort size effects, relative income, and marriage rates account for the paradoxical situation that when incentives for investments in intellectual competencies for men are low they are high for women. In other words, when marriage is delayed, female college enrollments tend to accelerate because women have higher incentives for investing in schooling.

These relationships played themselves out in Canada in the post-World War II period. The baby boom peaked in Canada in 1959 with a total fertility rate of 3.9 (Statistics Canada, 1984). When those born in the 1950-1960 period entered the labor market in the 1970s, the supply of younger workers exceeded demand, and relative incomes declined throughout the subsequent decade of the 1980s. During this period, for the first time in Canadian history, the number of young women entering and graduating from college with first degrees exceeded the number of young men (Statistics Canada, 1991, Table 2.3.1.2). In the United States, which has had a similar post-World War II demography to that of Canada, the 1980s also marked a period during which female college enrollments exceeded male college enrollments (National Center for Education Statistics, 1989, Table 148).

Today the Canadian demographic scene is one of decline. Total fertility rates are at 1.7 (Press, 1990, Table 3), well below the 2.1 replacement level. Newfoundland is somewhat anomalous, however, in that when Canada had its baby boom Newfoundland had what can only be described as a super boom. The total fertility ratio in Canada peaked in 1959 at 3.9, but in Newfoundland it peaked in 1961 with a rate of 5.3 (Press, 1986, Figure 1). Today, Canada's total fertility rate is at an all-time low, 1.7; in Newfoundland it is even lower, 1.5 (Press, 1990). In fact, Newfoundland's total fertility rate has dropped from the highest of any province in Canada to the lowest in just 12 years. Today, the population of Newfoundland is in absolute decline.

As a result of the high fertility in Newfoundland between 1957 and 1963, these babies, now adults, have flooded the labor market, resulting in low relative incomes and high unemployment rates (in excess of 20%). Hence there are low marriage rates and declining incentives for young men to invest in schooling. In contrast, the low marriage rates provide opportunities for young women to complete high school and to attend postsecondary institutions, that is, to increase their educational investments. One would predict, then, that the declining incentives for young men to invest in schooling

concomitantly with the increasing opportunities for women to invest might be expressed as differential achievements—higher achievements for girls than for boys.

The Model

In the light of the above discussion and observations, it was hypothesized that the gender differences in school achievement in Newfoundland, which are at odds with those in other educational jurisdictions in both Canada and the United States, could be explained in terms of unusually volatile demographic shifts. If, as is argued above, changes in relative incomes affect incentives to invest in intellectual competencies, it might reasonably follow that when conditions favor investments for girls, but not for boys, perceptions of the quality of school life would prove more positive for girls than boys. Similarly, one might conclude that perceptions of the quality of school life would govern both satisfaction and dissatisfaction with schooling. In this way quality of school life would mediate the effects of gender on student well-being. There is also the outstanding question of whether the perceived quality of school life will also mediate the gender-school achievement relationships.

Thus far, three basic hypotheses have been derived from the above interpretation of Easterlin's hypothesis, namely, that under societal conditions where relative incomes are low: (a) gender effects on the perceived quality of school life will favor girls; (b) gender effects on school achievement will be mediated by the quality of school life; and, (c) gender effects on overall student well-being will be mediated by the quality of school life. The question arises whether the hypotheses account for both cognitive and affective outcome differences at all grade levels. Since the developmental acceleration of girls throughout childhood seems to apply to both intellectual as well as physical development (Anastasi, 1958), girls tend to be in advance of boys at every age from kindergarten through puberty or early adolescence. Since this presumed developmental advantage ends at puberty, the maturity hypothesis would no longer be applicable from approximately grade 8 through the remainder of the high school grades. For this reason the sample chosen for this study was a grade 10 sample of mostly 15-16-year-olds.

But by this grade level the average age of boys is higher than the average age for girls due largely to three factors: early school leaving, grade repeating, and the special education selection process. The data for the 1988-1989 school year show that 24% of the students who were in grade 8 four years earlier (1984-1985) had withdrawn from school early and that of these 63% were boys (Department of Education, 1989, Table 6). In the 1989-1990 school year 1,779 students in grades 1 through 9 repeated a grade (Department of Education, 1989, Table 17). This was 2% of the student body and since grade repeating occurs every year, as much as 18% of the school population repeats at least one grade. Again, about 63% of grade repeaters are boys (Department of Education, 1989, Table 18). Similar findings exist with respect to special education placement. May, Boone and Hopkins (1988) report that 66% of the special education students in the province are boys. It is largely for these reasons that boys are underrepresented in the senior high school grades (10-12) and are on average older than the girls. Consequently,

the analytical models called for by the three hypotheses include age as a covariate or exogenous variable. Two additional control variables were also considered—socioeconomic status and urban-rural location of the school attended by the sample respondents. Socioeconomic status was included because the resources of families are powerful determinants of educational behaviors (Hauser, 1971). Urban-rural location was included because rural students in Newfoundland consistently fall behind their urban peers in terms of school achievement (Department of Education, 1991, Table 28 and Figure 12). Thus, following the Simon-Blalock approach to causal modeling (Blalock, 1962; Simon, 1952), the variables socioeconomic status and urban-rural location were added to the exogenous variable set. The full model that was tested is illustrated in Figure 1.

Instrumentation

Data were obtained from a test battery, a student questionnaire, a parent questionnaire, and school records. The reading comprehension and mathematics scores at the end of the grade 10 year were obtained from teacher administered subtests of the Canadian Tests of Basic Skills, High School Battery (CTBS, King-Shaw, 1989). The quality of school life instrument was a version of the instrument developed by Williams and Batten (1981) and which was modified by Roberts and Clifton (1991). This instrument includes 29 items designed to measure four dimensions of the quality of school life—opportunity to learn in school (8 items), adventure in learning or intrinsic interest in schooling (7 items), perception of status as a student (7 items), and students' perception of their teachers (7 items). In addition, the quality of school life instrument contained seven schooling satisfaction items

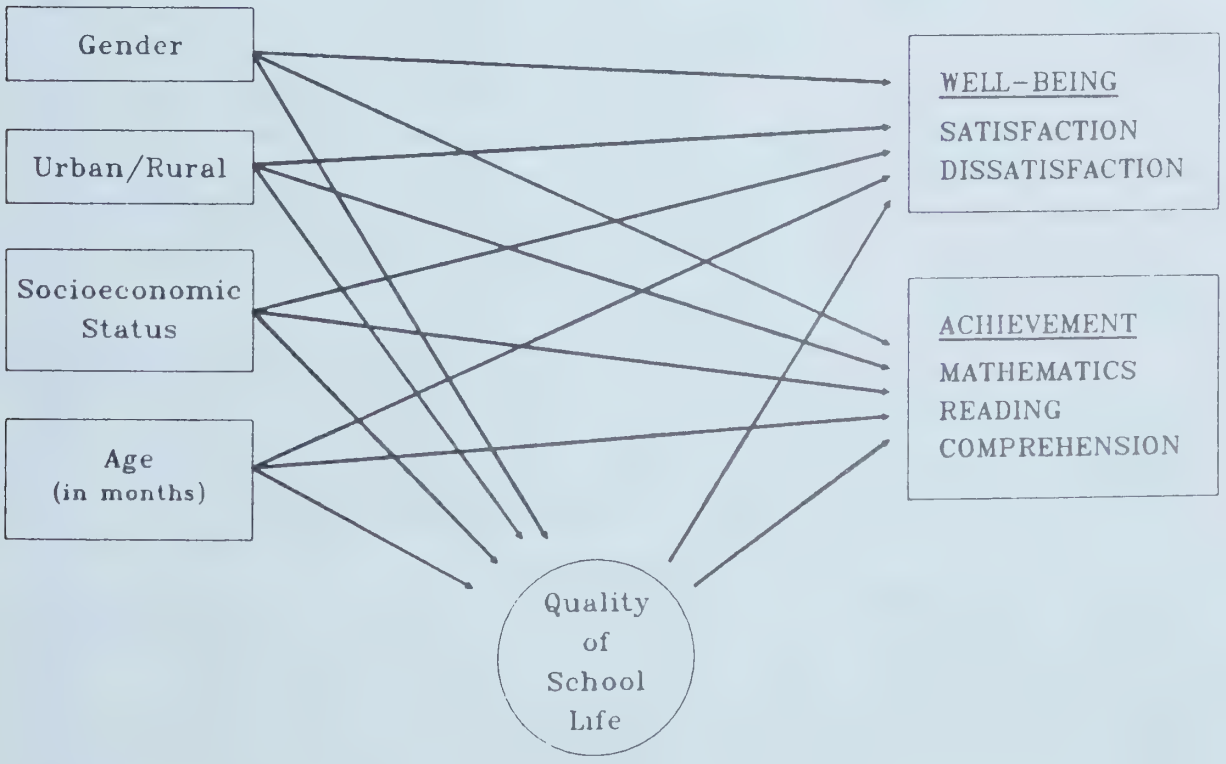

Figure 1. Conceptual model.

measuring positive affect and six schooling dissatisfaction items measuring negative affect.

Information on whether a school was located in an urban or rural community was obtained from Department of Education records. Communities with fewer than 5,000 people in the 1986 semi-decennial census were classified as rural. Information on parental education and occupation were obtained from a questionnaire completed by parents. Finally, the age and gender of the students were obtained from school records.

Sample

The population of schools with grade 10 classrooms was stratified by community size (urban and rural) and by denomination (Integrated and Catholic). Schools were chosen randomly from each cell. Schools located within the two school boards that refused to participate were replaced by "reserve" schools from the corresponding category. The sampled schools enrolled 450 grade 10 students (4.4% of the Newfoundland grade 10 population). Of these students, 317 (70.4%) received the informed consent of their parents to participate. Of those 317,182 (57.4%) were girls.

Measurement

Principal component analyses were conducted in order to construct weighted composites for seven variables—socioeconomic status, opportunity, adventure, status, teacher, satisfaction, and dissatisfaction. Because it was hypothesized (Williams & Batten, 1981) that opportunity, adventure, status, and teacher were dimensions of the quality of school life, they were used as indicators in a second-order factor analysis. It was also hypothesized that satisfaction and dissatisfaction were the dimensions of the student well-being construct. However, even though a well-being measurement model was estimated, it was not used except to estimate concurrent validity. It is argued here that when first-order composites are correlated with a second-order factor composite, the correlation is in fact an estimate of concurrent validity. In this case the concurrent validity is a measure of the relationship between first-order composites as indicators and the second-order composite as the criterion variable. The alpha reliabilities and concurrent validities of the composite variables are presented in Table 1.

The internal consistency reliabilities of the reading and mathematics achievement variables were not included in Table 1. In unreported analyses of CTBS results for selected Newfoundland grade level populations, the internal consistencies have been found to be marginally higher than those reported in the CTBS manuals (.90 ±.02, King-Shaw, 1990).

Results

The research hypotheses are embedded in the conceptual model presented in Figure 1. Model parameters for the five equations were estimated using ordinary least-squares regression: (a) quality of school life (QSL) regressed on the four exogenous variables (gender, urban-rural school location, socioeconomic status, and age); (b) schooling satisfaction regressed on the exogenous variables plus quality of school life; (c) schooling dissatisfaction regressed on the exogenous variables plus quality of school life; (d) mathe-

matics achievement regressed on the exogenous variables plus quality of school life; and (e) reading comprehension regressed on the same five variables as in (b) through (d). Equation (a) results are presented in Table 2. The estimates for equations (b) and (c) are presented in Table 3 and the estimates for equations (d) and (e) are presented in Table 4.

Because the metric for most model variables has little common language meaning, unstandardized betas are not reported. The effect parameters are standardized partial regression coefficients or path coefficients.

Discussion

In extending the cohort size hypothesis we argue that because prevailing demographic conditions in Newfoundland favor low relative incomes for young adults, the motivation to invest in intellectual competencies will be higher for female students than for male students. All the contemporary statistical evidence from official government sources supports the gender differences argument. High school girls in Newfoundland are performing as well as or better than high school boys in all school subjects. Similar trends for mathematics and science achievements have been reported in the United States.

It would seem to follow, therefore, that because girls are benefiting more than boys from their schooling opportunities, they would perceive both the quality of their school lives and their overall well-being as students in more

Table 1
Some Psychometric Properties of the Measurement Models

Measurement Model	Mean	Standard Deviation	N Items	Alpha Reliability	Concurrent Validity
SES (Socioeconomic Status)	49.73	10.00	2	.72	–
OPP (Opportunity to Learn)	50.00	10.00	8	.82	.83
ADV (Adventure in Learning)	50.00	10.00	7	.75	.77
ST (Student Status)	50.00	10.00	7	.71	.73
TCH (Perception of Teachers)	50.00	10.00	7	.80	.80
QSL (Quality of School Life)	50.00	9.45	4	.80	.67
SAT (Schooling Satisfaction)	50.00	9.21	7	.86	*
DIS (Schooling Dissatisfaction)	50.01	9.55	6	.70	–.562**

*Schooling satisfaction was used as a criterion for the QSL composite.

**Represents the convergent validity between schooling satisfaction and schooling dissatisfaction.

Table 2
Correlations, Direct Effects and t-Values for the Effects of Exogenous Variables on Quality of School Life

Intervening Variable	Exogenous Variables	Correlation (r)	Direct Effect	t-Value[a]
QSL				
	AGE	.035	.065	1.141
	GENDER	.092	.110	1.931
	UR	−.024	−.002	−.041
	SES	.086	.107	1.838

Multiple R = .14715
R^2 = .02165
[a]The t-value is for the direct effects only.

Key: AGE = Age in Months
GENDER = Male (1), Female (2)
UR = Urban (2), Rural (1)
SES = Socioeconomic Status
QSL = Quality of School Life

favorable terms than boys. Models designed to validate these propositions were formulated and estimated. The results are presented in Tables 2, 3, and 4. The findings were equivocal.

First, gender differences in the perceived quality of student life reported in Table 2 were not significant when controlling for age, urban-rural residence, and socioeconomic status differences. Second, gender difference in terms of their direct effects on schooling satisfaction and schooling dissatisfaction were not significant either (Table 3). When the effects of the gender model of schooling satisfaction were decomposed following Alwin and Hauser (1975), we were able to show that the total effect of gender on schooling satisfaction (Table 3) as mediated by the perceived quality of school life was substantially higher than the direct effect (0.146 compared with .073). The question arises, then, as to whether the ratio of the estimator to its standard error follows the *t* distribution (Folmer, 1981; Sobel, 1982; and Bollen, 1987). If it does, then the statistical significance of total effects can be estimated by the classical method. The jury is still out on this question, although recent findings (Bollen & Stine, 1990) show that in moderately large samples the distribution of an estimator is close to that assumed by the classical method. And since the sample size in the present study is large (N=317) we are able to assume that a total effect parameter of 0.146 will generate (using the classical method) a t=2.619 ($p < .05$). The same cannot be concluded about the total effect of gender on schooling dissatisfaction (TE=−.028). We conclude, therefore, that in terms of overall schooling satisfaction girls are generally more satisfied than boys, whereas in terms of overall schooling dissatisfaction there are no discernible gender differences. There is support for the proposition that girls find their schooling experiences more satisfying than do boys. But there is no support for two other propositions, namely, that girls will report higher quality school lives and less schooling dissatisfaction

Table 3
Correlations, Direct Effects, Indirect Effects, Total Effects and t-Values for the Effects of Independent Variables on the Dimensions of Student Well-Being

Outcome Variables	Independent Variables	Correlation (r)	Direct Effect	Indirect Effect	Total Effect	t-Value[a]
Satisfaction						
	AGE	.010	−.004	.043	.039	−.096
	GENDER	.134	.073	.073	.146	1.713
	UR	−.022	−.005	−.002	−.007	−.117
	SES	−.057	.006	.071	.077	.131
	QSL	.669	.661	—	.661	15.595*
Multiple R = .67272 R^2 = .45255						
Dissatisfaction						
	AGE	−.139	−.117	−.036	−.153	−2.454*
	GENDER	−.001	.033	−.061	−.028	.701
	UR	.063	.042	.001	.043	.891
	SES	−.062	−.022	−.059	−.081	−.447
	QSL	−.556	−.552	—	−.552	−11.734*
Multiple R = .57183 R^2 = .32699						

[a]The t-value is for the direct effects only.

Key: AGE = Age in Months
GENDER = Male (1), Female (2)
UR = Urban (2), Rural (1)
SES = Socioeconomic Status
QSL = Quality of School Life

*$t \geq 2.0$ is significant at the $p \leq .05$ level.

than boys. That there was an absence of support for either of these latter propositions calls for an ex post facto explanation.

It is explained above that the sampling frame for this study was based on grade, not age. Age-based samples are troublesome for the school authorities to obtain because they cross several grade levels. To avoid refusals to participate, grade-based samples are the compromise solution. But grade is not a good proxy for age at the high school level. Early school withdrawal, grade repeating, and special education arrangements distort the age-gender distributions of grade 10 classrooms because about two thirds of the members of these categories are boys. Two important consequences follow. First, there will be more female than male students in grade 10 and, second, the average age of male students will be higher than the average age of female students. In the present sample the correlation between age and gender was –0.121 ($t=-2.213$, $p < .02$), showing that older students were more likely to be male than female. Because it is plausible to assume that the absent male students

Table 4
Correlations, Direct Effects, Indirect Effects, Total Effects and t-Values for the Effects of Independent Variables on Reading and Mathematics

Outcome Variables	Independent Variables	Correlation (r)	Direct Effect	Indirect Effect	Total Effect	t-Value[a]
Reading Comprehension						
	AGE	−.227	−.190	.013	−.177	−3.770*
	GENDER	.054	.040	.022	.062	.798
	UR	−.261	−.210	.002	−.210	−4.199*
	SES	.366	.285	.021	.306	5.550*
	QSL	.224	.198	—	.198	3.983*
Multiple	$R = .49894$					
	$R^2 = .19452$					
Mathematics						
	AGE	−.264	−.268	.011	−.257	−5.119*
	GENDER	−.139	−.170	.019	−.151	−3.281*
	UR	−.176	−.150	.000	−.150	−2.885*
	SES	.272	.172	.018	.190	3.228*
	QSL	.162	.169	—	.169	3.277*
Multiple	$R = .44105$					
	$R^2 = .24894$					

[a]The t-value is for the direct effects only.

*$t \geq 2.0$ is significant at the $p \leq .05$ level.

were absent at least partially because they were ambivalent about schooling, the fact that they were absent would likely bias the findings in favor of support for the null hypothesis.

This explanation for the equivocal findings of the gender-schooling satisfaction relationship may also account for the lack of support for the thesis that gender effects on school achievement will favor girls. As shown by Table 4 results, there were no gender effects on reading comprehension, and the effects of gender on mathematics achievement was in favor of grade 10 boys. Again, these results are congruent with a sample bias in which girls were overrepresented.

Other features of the Figure 1 model are worth noting. In terms of the well-being equations reported in Table 3, none of the control variables was significant. Whether grade 10 students are satisfied or dissatisfied with their school careers depends largely on their perceptions of the quality of their school lives. Those reporting a low quality of school life tended to be the least satisfied and the most dissatisfied with schooling. Since these models had a good fit, with R-square values greater than 0.3, they seem worth taking

seriously. We know that there are four dimensions to the quality of school life construct and that school authorities might be able to manipulate these dimensions (opportunity to learn, adventure in learning, student status, and student identity) for the better, thereby bringing about enhanced student well-being. What is just as important, however, is that the quality of school life construct has significant positive effects on cognitive performance as measured by reading comprehension and mathematics ability. And these effects are over and above the effects of four control variables—age, gender, urban-rural residence, and socioeconomic status.

Note, however, that by grade 10 the debilitating effects of social background (SES) and the constraints of the rural environment on school achievement have already occurred—at least in Newfoundland. And note that, as predicted, age effects are both negative and statistically significant. The older students in grade 10 are the weakest students. In other words, by grade 10, age in the Newfoundland school system is to some extent a proxy for grade repeating and developmental delay.

Finally, we comment on the quality of student life model reported in Table 2. Our attempt to explain student perceptions of the quality of school life on the basis of selected demographic variables—age, gender, urban-rural residence, and socioeconomic status—was unsuccessful. These factors accounted for a mere 2% of the variance. Age, urban-rural residence, and socioeconomic status did not account for the quality of school life and, just as important, they did not account for schooling satisfaction either. Age, paradoxically, was a significant factor in accounting for schooling dissatisfaction (Table 3). Older students in grade 10, other things equal, were less dissatisfied with schooling than younger students, a finding for which we offer no explanation.

Conclusion

This article focuses on gender differences in student achievement in a province where girls are performing as well as and mostly better than boys in all school subjects at the high school level. Two theories purporting to explain gender differences in achievement are reviewed, namely, organic or endogenous psychological dispositions on the one hand, and cultural or social influences on the other. Neither seemed capable of accounting for the distinctive Newfoundland data. Instead, a cohort size hypothesis was examined that at the level of descriptive analysis seemed to provide a plausible explanation for gender differences in achievement in favor of girls. It generated the hypothesis that when relative incomes are low, male incentives to invest in intellectual competencies are low while, in contrast, when relative incomes are low, female incentives and opportunities are high. It followed from this that, if demographic circumstances were favorable for girls but not for boys, then the relationship between gender and both the perceived quality of student life and student well-being would favor girls.

These are dynamic relationships in that they are responsive over time to changing cohort size and relative incomes. Macro data suitable for the estimation of dynamic time series models were not available. Instead, static models based on cross-sectional data had to be used as the best substitute.

Not surprisingly, therefore, these exploratory micro-models with students as the units of analysis tended to be conjectural and the results somewhat equivocal. Nevertheless, total effect parameters in one model supported the thesis. That stronger support was not forthcoming may have been caused by unavoidable sampling design flaws that involved using grade level rather than age as the sampling frame. This limitation of the design generated a biased sample even though every effort was made to use appropriate statistical controls.

These limitations indicate the problems in testing the cohort size hypothesis as an explanation of gender differences in school achievement. The focus on gender differences, however, should not detract from some of the other interesting relationships found between student perceptions of the quality of their school lives and their overall student well-being. It was argued that the dimensions of the quality of school life construct were potentially policy manageable. These dimensions proved not only powerful predictors of well-being (schooling satisfaction and dissatisfaction), but also powerful predictors of literacy (reading) and numeracy (mathematics). School authorities might note, therefore, that increased emphasis on schooling quality could have important payoffs in terms of both schooling satisfaction and achievement for high school students.

Acknowledgment

This study was funded by the Department of the Secretary of State of Canada. The research assistants on the project were Shelley Hasinoff and Marguerite Baker. Jill Andrews and John Humber, Division of Evaluation and Research, Department of Education, Government of Newfoundland provided additional assistance.

References

Alwin, D.F., & Hauser, R.M. (1975). The decomposition of effects in path analysis. *American Sociological Review, 40*, 37-47.

Anastasi, A. (1958). *Differential psychology: Individual and group differences in behavior* (3rd ed.). New York: Macmillan.

Blalock, H.M. (1962). Four-variable causal models and partial correlations. *American Journal of Sociology, 68*, 182-194.

Bollen, K.A. (1987). Total, direct, and indirect effects in structural equation models. In C.C. Clogg (Ed.), *Sociological methodology* (pp. 37-69). Washington, DC: American Sociological Assosication.

Bollen, K.A., & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. In C.C. Clogg (Ed.), *Sociological methodology* (pp. 115-140). Oxford: Blackwell.

Bulcock J.W., & Mendoza, L. (1988). *The academic standing of the 1987 education graduates at Memorial University*. St. John's, NF: Institute for Educational Research and Development, Memorial University of Newfoundland.

Department of Education. (1988). *Education statistics*. St. John's, NF: Division of Evaluation and Research, Government of Newfoundland and Labrador.

Department of Education. (1989). *Education statistics*. St. John's, NF: Division of Evaluation and Research, Government of Newfoundland and Labrador.

Department of Education. (1990). *Education statistics: Elementary-secondary*. St. John's, NF: Division of Evaluation and Research, Government of Newfoundland and Labrador.

Department of Education. (1991). *Education statistics*. St. John's NF: Division of Evaluation and Research, Government of Newfoundland and Labrador.

Dimond, S.J., & Beaumont, J.E. (Eds.). (1974). *Hemisphere function in the human brain*. New York: Wiley.

Easterlin, R.A. (1980). *Birth and fortune*. New York: Basic Books.

Easterlin, R.A. (1987a). Easterlin hypothesis. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *The new Palgrave: A dictionary of economics* (vol. 2, pp. 1-4). London: Macmillan.

Easterlin, R.A. (1987b). *Birth and fortune: The impact of numbers on personal welfare* (2nd ed.). Chicago, IL: University of Chicago Press.

Ethington, C. (1990). Gender differences in mathematics: An international perspective. *Journal for Research in Mathematics, 21*, 74-80.

Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist, 43*, 95-103.

Ferguson, J. (1982). *Who turns the wheel?* Ottawa: Publications Office, Science Council of Canada.

Folmer, H. (1981). Measurement of the effects of regional policy instruments by means of linear structural equation models and panel data. *Environment and Planning, 13*, 1435-1438.

Galway, G., & Banfield H. (1991). *Profile '90: Educational indicators*. St. John's, NF: Department of Education, Government of Newfoundland and Labrador.

Geary, D. (1989). Comment: A model for representing gender differences in mathematical tasks. *Review of Educational Research, 59*, 185-213.

Hauser, R.M. (1971). *Socioeconomic background and educational performance*. Washington, DC: American Sociological Association.

Hollingworth, L.S., & Montague, H. (1914-1915). The comparative variability of the sexes at birth. *American Journal of Sociology, 20*, 335-370.

Hyde, J.S. (1981). How large are cognitive gender differences? A meta-analysis using w and d. *American Psychologist, 36*, 892-901.

Keeves, J.P. (1985). Sex differences in ability and achievement. In T. Husen & T.N. Postlewaite (Eds.), *International encyclopedia of education* (pp. 4438-4544). Oxford: Pergamon Press.

King-Shaw, E. (Ed.). (1989). *Canadian tests of basic skills: Teacher's guide*. Scarborough, ON: Nelson Canada.

King-Shaw, E. (Ed.). (1990). *Canadian tests of basic skills: Manual for administrators, supervisors and counsellors*. Scarborough, ON: Nelson Canada.

Linn, M.C., & Hyde, J.S. (1989). Gender, mathematics and science. *Educational Researcher, 18*, 17-19, 22-27.

Maccoby, E.E. (Ed.). (1966). *The development of sex differences*. Stanford, CA: Stanford University Press.

Maccoby, E.E., & Jacklin, C.N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.

Marsh, H. (1989). Sex differences in the development of verbal and mathematical constructs: The high school and beyond study. *American Educational Research Journal, 26*, 191-225.

May, S., Boone, W., & Hopkins, B. (1988). *Sex-related differences in mathematics: Participation, achievement, and attitudes in the Newfoundland and Labrador setting*. St. John's, NF: Division of Evaluation and Research, Government of Newfoundland and Labrador.

National Center for Education Statistics. (1989). *Digest of education statistics*. Washington, DC: Department of Education, Office of Educational Research & Improvement.

Pearson, K. (1897). *The chances of death and other studies in evolution* (vol. 1). London: Arnold.

Press, H.L. (1986). *Trends in education: Demographic projections*. St. John's, NF: Division of Evaluation and Research, Government of Newfoundland and Labrador.

Press, H.L. (1990). *Toward 2000: Trends report II, elementary-secondary projections*. St. John's, NF: Division of Evaluation and Research, Government of Newfoundland Labrador.

Resnick, S.M., Berenbaum, S.A., Gottersman, I.I., & Bouchard, T.J. (1986). Early hormonal influences on cognitive functioning in congenital adrenal hyperplasia. *Developmental Psychology, 22*, 191-198.

Roberts, L.W., & Clifton, R.A. (1991). *Measuring the quality of university student life*. Winnipeg, MB: Department of Sociology, University of Manitoba.

Sadker, M., Sadker, L., & Klein, S. (1991). The issue of gender in elementary and secondary education. In G. Grant (Ed.), *Review of Research in Education* (vol. 17, pp. 269-334). Washington, DC: American Educational Research Association.

Sarbin, T.R. (1954). Role theory. In G. Lindzey (Ed.), *Handbook of social psychology* (vol.1, pp. 223-258). London: Addison-Wesley.

Seward, J.P., & Seward, G.H. (1980). *Sex differences: Mental and temperamental*. Lexington, MA: C.D. Heath.

Simon, H.A. (1952). On the definition of the causal relation. *Journal of Philosophy, 49*, 517-528.

Sobel, M.E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 290-312). San Francisco, CA: Jossey Bass.

Springer, S.P., & Deutsch, G. (1981). *Left brain, right brain*. San Francisco, CA: Freeman.

Statistics Canada. (1984). *Fertility in Canada: From baby boom to baby bust*. Ottawa: Minister of Supply & Services, Government of Canada.

Statistics Canada (1991). *A statistical portrait of university education in Canada*. Toronto, ON: Council of Ministers of Education Canada, Government of Canada.

Tyler, L.E. (1956). The psychology of human differences (2nd ed.). New York: Appleton-Century-Crofts.

Tyler, L.E. (1969). Sex differences. In R.L. Ebel (Ed.), *Encyclopedia of educational research* (4th ed., pp. 1217-1221). Toronto, ON: Macmillan.

Wheatley, G.H., Mitchell, R., Frankland, R.L., & Kraft, R. (1978). Hemisphere specialization and cognitive development: Implications for mathematics education. *Journal for Research in Mathematics Education, 9*, 20-32.

Williams, T., & Batten, M. (1981). *The quality of school life*. Hawthorn, Victoria: Australian Council for Educational Research.

The Alberta Journal of Educational Research *Vol. XXXVII, No. 3, September 1991, 225-239*

GAALEN ERICKSON
University of British Columbia

and

SANDRA FARKAS
Vancouver Technical Institute

Prior Experience and Gender Differences in Science Achievement

A number of conjectures voiced in the literature concerning the role of prior experiences as a possible factor in accounting for the observed gender differences in science achievement tests were examined. A set of 23 multiple-choice items that displayed significant gender differences on a previous British Columbia science assessment test were administered to 238 grade 12 students. Fifteen of these students were selected for a follow-up interview to probe the reasons they gave for their responses to these items. In particular, they were encouraged to identify any previous experiences they might have had that might have enabled them to respond to the item. Four categories were generated from their explanations of why they responded as they did to the items. These were school-based experiences, informal experiences, common knowledge, and a final miscellaneous category. The results indicate that females consistently appealed to school-based experiences much more than did the males (even though this did not translate into better results on the items). A second finding was that many of the items elicited spontaneous negative reactions for many of the females and virtually none of the males. In examining these results and drawing on the rapidly expanding literature in this field of inquiry, it was concluded that efforts to remove the gender bias in achievement testing will require a systematic, empirical analysis of the experiential contexts of items.

Introduction

That the issue of gender effects in science achievement and enrollment patterns remains one of current interest can be readily inferred from the continued expansion of the research literature in this area (Bearlin, Annice, & Elvin, 1988; Kelly, 1987a; Koshland, 1988; Linn & Hyde, 1989). An even more important indicator of the vitality of the field in science education is the increasing institutionalization of resources aimed at developing effective policies and practices to address some of the problems of females and science that have been identified in the literature. National and international organizations have been created to sponsor conferences, workshops, and projects designed to create and disseminate materials and strategies for the

---

Gaalen Erickson is an associate professor in the Department of Mathematics and Science Education, and Associate Director of the Centre for the Study of Teacher Education. His research interests are in the areas of science education, gender issues in the assessment context, and teacher education.

Sandra Farkas is a teacher in the Vancouver School District. Her research interests are in gender equity and classroom practice.

teaching of science/technology subjects in a way that will not disadvantage females. Of the international organizations, the Girls and Science and Technology (GASAT) group has been the most active, having sponsored five international meetings in the past eight years.

On a smaller scale many action groups at provincial/state or even at local levels (universities or school districts) have been established for the purposes of developing gender-conscious curricula, organizing workshops and inservice programs for educators, and lobbying political organizations. Examples of many of these action groups all over the world have been reported at recent GASAT meetings (Daniels & Kahle, 1987; Harding, 1986; Raat, 1981; Sjoberg, 1983).

Although the proliferation of these groups does signal a recognition of the importance of gender issues, it does not mean that all of the questions concerning gender effects have been resolved. One area of continuing discussion in both the research literature and in the various programs is that of the nature of the possible sources of the observed gender differences. While much of the earlier debate in the 1970s focused on the roles played by biological versus sociological factors (Kelly, 1981; Kimball, 1982; Maccoby & Jacklin, 1974; Sherman, 1978), most of the recent research and program development has been directed toward identifying those sociocultural factors that contribute to the observed differences, particularly those that might be amenable to change by altering current school practices (Bentley & Watts, 1986; Crossman, 1987; Haggerty, 1987; Linn & Hyde, 1989; Martin, 1989; McClintock Collective, 1988).

One of the arguments that has been consistently made to account for girls' lower achievement scores in the physical science subjects is that in their growing years girls tend to have fewer and different kinds of experiences with a wide range of physical phenomena relevant to these subjects (Johnson & Murphy, 1984; Kahle & Lakes, 1983; Lie & Bryhni, 1983; Sjoberg, 1989). In some instances these arguments are advanced as intervening premises suggesting that girls' prior experiences could account for observed gender differences in spatial and mechanical reasoning scores that may in turn be the basis for differences in achievement scores (Finn, Dulberg, & Reis, 1979; Maccoby & Jacklin, 1974). In other cases the argument is constructed to implicate prior experience directly in accounting for the differences in standardized achievement tests or enrollment patterns (Erickson & Erickson, 1984; Johnson & Murphy, 1984; Kelly, 1987b; Parsons-Chatman, 1987). For example, Erickson and Erickson (1984) examined a number of items in a large-scale achievement test that displayed substantial gender differences and proposed that some of the observed differences could be explained on the basis of the item context. Those items that were embedded in a context that was much more familiar to males (i.e., were part of their "sphere of experience") may have conveyed an advantage to the males by allowing them to construct more plausible responses to the items even if they did not "know" the answer from their direct school experiences in science. For example, items that used balls, electrical circuits, compasses, and other materials or situations that are more likely to be a part of a boy's everyday experiences

seemed to have particularly large differences in achievement scores between males and females.

Bateson and Parsons-chatman (1989) went one step further in checking out the validity of this conjecture. They developed a scale, which they labeled *gender orientation,* to characterize items in terms of the degree to which they were associated with or related to the prior experiences of grade 7 girls or boys. They then asked 20 adults (including both elementary school teachers and parents) to use this scale to rate 120 items that were used in the 1986 British Columbia Science Assessment (Bateson, Anderson, McConnell, Rutherford, & Dale, 1986). Correlating the mean gender orientation rating for each item with the differences in the achievement scores between boys and girls on these same items, they were able to obtain an estimate of the relationship between these two variables. The correlation coefficient they reported for the entire 120-item test was .53 while the correlations for some of the subtests ranged from a low of .18 to a high of .84 (Bateson & Parsons-chatman, 1989). These data suggest that prior experience may indeed be important in influencing performance on achievement tests, but because Bateson and Parsons-chatman had to use judgments from adults to assign gender orientation scores, their empirical support for the prior experience argument is indirect. More direct investigations of the prior experience hypothesis, both to assess its validity and to begin sketching out its dimensions in qualitative terms, are clearly desirable.

Overview of the Study

The work reported here represents an initial endeavor to begin mapping out in more detail some of the relevant qualitative factors in this problem area. The study is clearly exploratory in nature and was designed to examine in a more interpersonal context the apparent impact of prior experience on students' performance on a multiple-choice science achievement test. The specific intent of the present study was to obtain more information about prior experience from the students themselves rather than using more indirect methods. The value of obtaining data directly from the participants regarding the meaning they attach to particular situations (such as responding to a multiple-choice item) has been underscored in the recent debates on competing traditions of educational inquiry (Erickson, 1986; Firestone, 1987; Howe, 1988; Marton, 1988; Smith & Heshusius, 1986). Since this study is aimed at uncovering some of the meaning that students attach to responding to science achievement items, it is located in the growing tradition of *interpretive research* (Soltis, 1984).

The Research Questions

Several specific questions were used to guide the collection and the analysis of the data for this study. These specific questions were:

1. What are the different kinds of explanations students give in accounting for their choices on a set of science achievement items? Can these explanations be categorized according to the type of prior experiences that they draw on?
2. Are there any gender differences in these explanatory categories?

3. Does the item context account for any gender differences?

Methods

Data Collection

Two related data bases were generated in the study. One was obtained from a sample of 238 grade 12 students selected from three schools in the Vancouver Public School system. These students were administered a 23-item, multiple-choice test (see Farkas, 1986, for a complete description of the items and the procedures employed). Table 1 provides a classification of the items used by goal area and content area. As shown, the items were distributed among the cells formed by crossing the goal and content areas. The 23 items had been used earlier in a larger, province-wide science assessment in 1982 (Taylor, Hunt, Sheppey, & Stronck, 1982) and had yielded differences in scores in favor of males ranging from 8 to 30 percentage points.

The items were administered to 238 students for three reasons. First, we wanted to ensure that there was some stability in the gender differences that had been observed in these items in the previous (1982) assessment. Second, we wanted to simulate as closely as possible the types of conditions (e.g., large group, limited time, and not in a science class) under which the BC Assessment measures are normally administered. Finally, we selected from this sample of students a smaller group to interview. We thought it was important that the students we interviewed had originally responded to the items under assessment-like conditions.

The test was administered by the regular classroom teachers to the students in their English or consumer education classes. (These were the only two compulsory grade 12 subjects at the time of the study.) Using these students enabled us to obtain a much more heterogeneous and representative sample than if we had used grade 12 science classes.

Table 1
Classification of Test Item Numbers by Goal and Content Area

	Content Area			
Goal Area	Biology	Chemistry	Physics	Earth/Space
I. Science Processes				
a. interpret data	20,21[a]		1,2	
b. identify and control variables		11		
II. Knowledge, Recall				
a. understanding	9	18	6	5,16
b. safety procedures		7,8,17,22	13	
III. Higher Level Thinking				
a. evaluate evidence for conclusions			15,23	12
b. solve abstract problems		10	14,19	3,4

[a]Item numbers

For our purposes the most important data for the study were those obtained from a series of semistructured personal interviews with 15 students (8 males and 7 females) selected from the larger group. These students were selected by their respective teachers on the grounds that they represented a typical cross-section of students in terms of academic performance and other background variables. The interviews, which were conducted by the second author within a week after the students had completed the original test, were designed to identify linkages between the students' self-reported prior experiences and their written responses to the multiple-choice items included in the test.

Each interview began with a brief orientation period developed to put the students at ease and to explain that we were interested in their ideas about how they arrived at their answers to the items and not whether they had the "right" answer. The students were then given their own unmarked test so that they could review each question and their response to it. The students were asked to reread each question and their response to it. They were then asked some variation of the following question: "Could you remember what you were thinking when you answered this question or can you explain why you selected this answer?" The interviewer often asked for further elaboration of their response depending on the nature of their initial response. These interviews, which typically lasted for 30 to 40 minutes, were recorded on audiotape for later transcription and analysis. Illustrative comments from some of these interviews are provided in the Results section.

Data Analysis

The data analysis of the student responses to the 23-item test consisted of scoring the items according to correct and incorrect answers. Analysis of the interview data focused on developing an appropriate set of categories to account for the explanations offered by the students in justifying their responses to a given item. First, a preliminary list of all the reasons offered by the students was prepared. These reasons, many of which were based on students' recollections of previous school experiences, were then grouped into four categories: school-based experiences, informal experiences, common knowledge, and miscellaneous. Each of these categories is described and illustrated in the presentation of results of the analysis of the students' responses.

Results

Written Tests

The percentage correct between the males and females and the difference in these percentages for both the 1982 assessment and the current administration are presented in Table 2 for each of the 23 multiple-choice items included on the test.

There were three items (8, 15, and 22) for which the difference scores varied by more than 10% between the two assessments. We have no easy explanation for these changes except to say that they were created from a sizable decrease in the performance by the males combined with a small increase in performance by the females. It is interesting to note in passing that items 8 and 22 assessed chemistry safety procedures. The issue of safety

Table 2
Percentage Scores for Selected Science Achievement Items From 1982 BC Science Assessment and Current Study by Gender

	1982 Results (N = 22,110)			Current Results (N = 238)		
Item No.	Male	Female	(M - F)	Male	Female	(M - F)
1	46%	38%	8%	49%	32%	17%
2	43	26	17	42	27	15
3	48	26	22	60	31	29
4	41	19	22	42	26	16
5	36	25	11	54	36	18
6	60	34	26	65	38	27
7	53	38	15	55	36	19
8	27	17	10	20	22	–2
9	71	57	14	72	48	24
10	50	32	18	57	32	25
11	57	46	11	68	52	16
12	52	38	14	52	45	7
13	74	44	30	66	35	31
14	76	55	21	72	51	21
15	69	44	25	54	47	7
16	58	39	19	57	39	18
17	80	67	13	67	51	16
18	72	62	10	73	61	12
19	69	46	23	70	50	20
20	47	37	10	41	32	9
21	66	58	8	58	40	18
22	70	46	24	60	51	9
23	60	45	15	63	51	12
Mean	57.6	40.8	16.8	57.3	40.6	16.7
SD	13.7	13.0	6.2	12.3	10.0	7.6

Note: Percentage scores are rounded off to the nearest whole number.

has recently received considerable attention in science classes, but this emphasis does not seem to be reflected in the males' performance on these two items. But in aggregate scores, the 1982 and current samples were virtually identical for both the item scores and the difference scores.

However, given the similarity of both the mean scores and the difference scores for both groups, we would argue that the gender differences in the responses to the items have remained relatively constant over the four-year interval between the two test administrations.

Interviews

The interview data suggest that the students appeared to be comfortable in discussing the reasoning and/or experiences that they used in selecting their responses. A wide range of comments and recounting of experiences was elicited from these interviews. An illustration of the type of comments made is provided in the following excerpt (I = Interviewer, S = 17-year-old female student):

> I: What were you thinking when you chose the answer you did for question 7? [A question on the proper procedure for lighting a Bunsen burner.]
>
> S: Okay. When I was in grade 9 I had a bad experience with a Bunsen burner. I don't know whether my answer is right or not. When I lit the Bunsen burner it kind of shocked me. My science teacher gave me a sheet and said that you should read it and study it and know where the air control should be. Well, I'm never going to light another Bunsen burner; I'm going to get someone else to do it so I never read it. So, what I put for my answer was number four because air allows a flame to burn. I didn't want to put "I don't know" because I kind of had a bit of knowledge about it.

Four explanatory categories were constructed from the student responses. The first, labeled school-based experiences, included any student comments that referred to some experience related to, or knowledge obtained from, some previous school activity. The excerpt given above, for example, was placed in this category. A second example is the comment: "Well, it's something I remember my teacher telling me to do—you always mix the water with the acid" that was provided by a female responding to an item asking for the appropriate laboratory procedure for diluting acid. It should be noted at the outset that as in the foregoing example an appeal to some school-based experience did not necessarily result in a "correct" answer.

The second category, informal experiences, was used to categorize student references to experiences or events such as hobbies, playing games, television, radio, and recreational reading. In other words, this category included experiences obtained outside the school context. For example, a male responded to a question on the purpose of circuit breakers by stating: "That I know from stuff around the house. Fuses blow in our house all the time. I usually go downstairs and switch the breakers."

Common knowledge was the third category, and it included references to knowledge that the student held where its source could not be readily identified. A number of students made statements such as "Well, that's just basic knowledge. Something I learned a long time ago" and "This sort of made sense to me. It was logical."

The final category, which we have simply labeled as miscellaneous, included such things as guessing at the answer, not being able to remember how they had obtained an answer, or working out the answer by a process of elimination. The types of comments included in this category are typified by the following female response: "Oh boy, I don't know what would happen. I have too much pride to put 'I don't know,' so I guessed." A slightly different

rationale was offered by this male student: "At the time I knew the answer, but it seems too long and complicated to remember what I was thinking."

In response to our first research question, then, we think these examples demonstrate that a semistructured interview does appear to be an effective means of identifying a wide range of prior experiences that the students claimed to have used in generating their responses to the written test. Furthermore, it was possible to develop a category system that enabled us not only to accommodate all of the students' responses, but, more importantly, provide some potentially fruitful conjectures about how males are able to draw on a wider repertoire of experiences in answering selected science questions.

The second question we examined was whether there were any gender differences in the responses provided by the students. One way of identifying these differences was to tabulate these explanatory responses by category and by gender. Table 3 provides these data. An inspection of this table reveals that the response category with the largest gender difference is that of school-based experiences. The females referred to such experiences for just over one half of their responses, while the males referred to school-based experiences for just over one third of their replies. In contrast, males more than females referred to other experiences as a basis for their responses. In brief, it seems clear that there were indeed consequential gender effects among the types of prior experiences offered by the students. The females more than the males relied more on information/experiences remembered from their school science, whereas the males appealed more than females to knowledge obtained from their out-of-school interests and what we have termed as common knowledge.

The final research question of this study was addressed in two ways. The first was to identify items where the context is most likely to be gender sensitive (Bateson & Parsons-chatman, 1989; Erickson & Erickson, 1984) and examine the students' explanatory responses for those items. The second approach was to identify those items that elicited a spontaneous negative reaction from the students and to determine the source of these reactions. It is noteworthy that all the students, with a single exception, who responded in this spontaneous manner to particular features in the question stems were females.

Table 3
Male and Female Explanatory Responses by Experience Category

	(n)	School-based Experiences	Informal Experiences	Common Sense	Misc.	Total No. of Responses
Male	8	66 (36)	38 (21)	40 (22)	40 (22)	184
Female	7	87 (54)	26 (16)	22 (14)	26 (16)	161
Difference (M - F)		(–18)	(5)	(8)	(6)	

Note: Percentage of explanations for all 23 items is given in brackets and rounded off to the nearest whole number.

A context that has been identified by most researchers in this field as being very gender sensitive is that of electricity (Johnson & Murphy, 1984; Kelly, 1987b; Parsons-Chatman, 1987). We found this was apparent in our study as well. Question 13, which generated the largest difference score (31%), asked about the purpose of a fuse or circuit breaker. In describing the reasons for their answers, six out of the seven females interviewed referred to some type of school-related experience to justify their selection. As indicated earlier, however, this attempt to remember relevant ideas or details from earlier school experiences often resulted in incorrect answers or confusion about the concept being assessed as illustrated by the following response:

> I know what a circuit breaker is. It makes things go in two directions. We did it in grade 10. I wasn't into it. I didn't put it in my long-term memory. It was a little tedious and hard for me to understand.

It seems that she may have confused a circuit breaker with some sort of switch used in a laboratory exercise on electrical circuits. More importantly her comment reveals a noticeable lack of interest in the topic.

On the other hand, only three out of the eight males interviewed used the school experience category for this item; the other five males referred to a range of informal experiences such as:

> I built a basement suite downstairs and I took stagecraft as well.

> I'm interested in electronics so I know about circuit breakers.

Items related to the wearing of seat belts (item 6), the stopping distances of cars traveling at different speeds (items 1 and 2), or identifying the type of substances that must be kept away from an open flame (item 17) yielded similar results in terms of the males choosing to draw on their informal and everyday experiences more than did the females. It would appear that on many of the items males had some advantage in answering them correctly by virtue of their ability to draw on a wider range of out-of-school experiences.

A second source of data used to address the issue of the role of the item context on student responses was the spontaneous, affective comments (which typically were negative) made by the female students as each question was discussed. Close to half of the items elicited an immediate negative response from more than one of the females interviewed, whereas only one male responded negatively and that was to only one item. Typical of the negative spontaneous female responses to an item is the following: Shown a picture of a Bunsen burner (item 7) and asked about the proper procedure for lighting it, several females made comments like:

> Bunsen burners scare me. I never did learn how to use them.

> Well, I got my partner to do it. I went and did the other stuff. Maybe I was afraid of Bunsen burners.

Similar negative comments were frequently voiced by the females as they were asked to explain their thinking on other items involving physical science concepts like circuit diagrams (item 19) or astronomy concepts (items 3 and 16):

> Oh no, not another physics item!
>
> I remember doing a thing on this but most of this comes from talking to the guys. I'm not into technical stuff.

These negative kinds of reactions may have blocked any serious cognitive engagement with the item and resulted in the females simply guessing at a response rather than trying to reason or even to recall some knowledge that might be of some assistance to them.

While our presentation of the results on the role of item context in creating (or at least exacerbating) gender differences has been brief, we think such results are indeed suggestive of an important factor that has frequently been overlooked in the construction of achievement tests.

Discussion

The data that we have sketched out above are consonant with other recent commentators in this field (Bateson & Parsons-chatman, 1989; Kelly, 1987a; Linn & Hyde, 1989; Sjoberg, 1989). Our results suggest that males seem to enjoy an advantage when it comes to retrieving knowledge or experiences that might be helpful in answering science achievement items. This advantage appears to be operative in two distinct but related ways. First, it seems that many of the items were embedded in contexts that allowed the males to utilize their prior everyday or "informal experiences"—contexts that typically did not seem to be accessible to the females in our study. Thus the males talked about their knowledge obtained from activities like wiring a basement suite, experiences with driving and fixing cars, using a compass in an orienteering course, or messing around with chemistry sets and model rockets. This is not to say that females do not have any opportunities to become involved in activities like this, but that on balance these sorts of activities tend to be more associated with the socialization experiences that are part of being brought up as a male in our society (Fischer, 1982; Johnson & Murphy, 1984; Lie & Bryhni, 1983). Furthermore, our interview results indicated that for items that involve topics like electricity, braking distances for cars, and appropriate safety procedures for working with chemicals, the females endeavored to draw on previous school experiences—experiences that often did not make much sense to them at the time when taught, let alone now. It is not surprising, then, that even though the females attempted to utilize their school knowledge on slightly more than half of the items (see Table 3), this strategy was not particularly successful for many of them and they still, on average, performed less well on these items than the males.

This finding, although different in origin, bears some similarity to the point made by Linn and Hyde (1989) about how the use of school-based knowledge and techniques may actually hinder female performance in mathematics achievement testing. They argue that teachers often emphasize the importance of going through all the steps of a mathematical algorithm or procedure and showing one's work. However, this is very time-consuming when solving a large number of problems on a timed achievement test. The end result is that those students who use these school-learned methods often do not complete the test. And the evidence seems to suggest that females are

more likely to resort to the techniques that they learned in school than are males.

It is interesting to note that in those situations where the females were able to make some connections to their life experiences, they were more successful in answering the questions. For example, on a question concerning the molecular motion of steam, a female explained how she arrived at her correct response by stating: "What I thought about steam is that it moves about a room. When you put a kettle on the stove and close the kitchen door the windows fog up." Or, in another question on the increase of water pressure with depth, several females indicated that they knew the answer because of their experience with diving. The more general point to be made, then, is that many of the items used to assess the understanding of physical science concepts confer an advantage to males because the item contexts enable them to deploy more effective answering strategies.

Another finding of general interest in our study was that a number of the items elicited spontaneous negative reactions by a number of the females. Many of the items referred to objects or situations where the females had frightening experiences (such as with a Bunsen burner), where they lacked confidence and experience (the electricity items), or which the females construed as boring and uninteresting ("not another physics item"). The result tended to be that the female students approached such items tentatively or with disinterest. On the other hand, they appeared to show more enthusiasm for items that made more intuitive sense to them.

These issues and results could be interpreted in the more narrow sense as suggesting we simply ensure that we develop assessment questions that are gender neutral. While we would certainly encourage such endeavors, we also agree with the critics who argue that the assessment issue is only a very small symptom of a much larger set of educational and sociocultural problems. We must ultimately reconceptualize our basic ideas about the nature of scientific inquiry (Easlea, 1981; Keller, 1985; Martin, 1988, 1989), and in particular our ideas about the design and delivery of science curriculum materials and instructional strategies (Bentley & Watts, 1986; Lewis & Davies, 1988; Manthorpe, 1982; McClintock Collective, 1988; Smail, 1987).

Our study points to some interesting directions for initiating some aspects of this reconceptualization process. One such direction entails a more careful examination of the subtle yet pervasive ways that prior experience and the accompanying role expectations constructed from these experiences influence both students and teachers in their daily interactions in the classroom. The importance of this influence of prior knowledge on subsequent learning activities has been the subject of intense scrutiny in the area of science and mathematics concept learning (Cobb, 1988; Driver, 1989; Novak, 1987; von Glasersfeld, 1989; White, 1988). The methods and findings from the extensive literature generated by this constructivist frame for educational inquiry would seem to have significant implications for extending the scope and shifting the nature of the research agenda in the field of gender and science (Vegting, 1987).

A brief example will illustrate what this shift in perspective would involve. Some researchers have argued that the number and kind of science courses

taken by males in comparison with females is an important factor in accounting for the differential performance in the physical sciences (Kahle, 1985; Kelly, 1981; Linn & Hyde, 1989). A more complex version of this argument has been advanced by Klainin and Fensham (1987) and Klainin, Fensham, and West (1989b) as a partial explanation of the results they obtained from a Thailand study where the females outperformed the males in several tests of chemistry achievement and did equally as well in a set of physics tests. The fact that the Thailand educational system required all students enrolled in the science curricular stream to study chemistry, physics, and biology at the secondary level is no doubt an important factor. However, given the implications that we have drawn from our study, we would suggest that their reporting of a different set of cultural norms and expectations regarding gender and science (e.g., the perception by the students that chemistry activities are "women's work" and that over 90% of the chemistry teachers in their study were female) represents a very provocative and interesting finding. These differing norms and attitudes seem to lead to behavioral outcomes on the part of the female students that has not been reported in the Western literature (e.g., the females in the chemistry classes were found to be "more dominant in classroom exchanges than the boys" (Klainin & Fensham, 1987).

Rather than attempting to quantify and classify school experiences in terms of the science courses taken, a constructivist perspective would focus more on an examination of the nature of the curricular content and the instructional strategies so as to determine the extent to which they acknowledge the prior experiences and understandings of all students. Further, it would be more concerned with the quality of the interactions between teachers and students and endeavor to illustrate the importance of continually checking out the meaning that students, as well as teachers, confer on classroom events and discourse. As a case in point, Erickson and MacKinnon (1991) recently documented in some detail an instance of a teacher who followed up on a simple comment made by three females working together on a laboratory problem involving electrical circuits. Initially the teacher thought that the students had the "correct" answer to the problem and that they understood the underlying concepts. But he was uneasy about one of their diagrams, and with some probing, combined with considerable patience and encouragement, the students discovered that their earlier understanding was quite confused and inadequate. Through additional experiments and considerable discussion and debate among themselves they were able to construct a more complete and functional understanding of electrical circuits. This short teaching encounter proved to be extremely satisfying for the students and, as a result, the teacher.

One of the important messages of these types of in-depth, classroom-based case studies is that they reveal to practitioners as well as researchers not only the inherent complexity of much of what we attempt to teach students but, more importantly, the value of listening carefully and checking out the meanings that students often attribute to their classroom experiences. This advice would appear to be particularly relevant when interacting with females, given the evidence presented above and the documented ten-

dencies of females to conform to school procedures and to produce "school answers" (Haggerty, 1987; Linn & Hyde, 1989).

In conclusion, we suggest that the present study represents a useful starting point for further explorations on gender and science. The qualitative data presented above raise a series of important and provocative questions. To what extent will females continue to rely on prior school experiences when in most instances this strategy did not appear to be fruitful in assisting them to clarify or answer assessment questions? Is this response pattern an example of females lacking confidence in their own personally constructed knowledge (Linn & Hyde, 1989), or is it a matter of the females being more sensitive to the implicit messages of school assessment exercises? That is, do females assume these exercises are measuring school learning and therefore attempt to provide appropriate school-based responses (Klainin, Fensham, & West, 1989b)? Is it possible to develop gender neutral assessment items as some researchers have encouraged? What is meant by gender neutrality as it pertains to achievement testing or curriculum development? Our results and the work of others (McLarty, Noble, & Huntley, 1988; Wedman & Stage, 1983) indicate that removing the gender bias in achievement testing needs to go well beyond the "superficial gender-related item modifications" tried by McLarty et al. (1988) to a more systematic examination of the experiential contexts of items.

References

Bateson, D., Anderson, J., McConnell, V., Rutherford, C., & Dale, T. (1986). *The 1986 British Columbia science assessment: General report*. Victoria, BC: Ministry of Education.

Bateson, D., & Parsons-Chatman, S. (1989). Sex-related differences in science achievement: A possible testing artifact. *International Journal of Science Education, 11*(4), 371-385.

Bearlin, M., Annice, C., & Elvin, A. (1988). *WASTE: Women and science education*. Canberra: Commonwealth Schools Commission.

Bentley, D., & Watts, M. (1986). Courting the positive virtues: A case for feminist science. *European Journal of Science Education, 8*, 121-134.

Crossman, M. (1987). Teachers' interactions with girls and boys in science lessons. In A. Kelly (Ed.), *Science for girls?* Milton Keynes, UK: Open University Press.

Cobb, P. (1988). The tension between theories of learning and theories of instruction in mathematics education. *Educational Psychologist, 23*, 87-104.

Daniels, J., & Kahle, J. (Eds.). (1987). *GASAT IV: Contributions from the fourth conference of Girls and Science and Technology*. Ann Arbor, MI: University of Michigan.

Driver, R. (1989). Changing conceptions. In P. Adey (Ed.), *Adolescent development and school science*. London: Falmer.

Easlea, B. (1981). *Science and sexual oppression*. London: Weidenfeld & Nicolson.

Erickson, F. (1986). Qualitative methods in research on teaching. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 119-161). New York: Macmillan.

Erickson, G., & Erickson, L. (1984). Females and science achievement: Evidence, explanations, and implications. *Science Education, 68*(2), 63-89.

Erickson, G., & MacKinnon, A. (1991). Seeing classrooms in new ways: On becoming a science teacher. In D. Schön (Ed.), *The reflective turn: Case studies of reflection in and on practice*. New York: Teachers College Press.

Farkas, S. (1986). *Gender differences in science tests*. Unpublished master's thesis, University of British Columbia.

Finn, J., Dulberg, L., & Reis, J. (1979). Sex differences in educational attainment: A cross-national perspective. *Harvard Educational Review, 49*(4), 477-503.

Firestone, W. (1987). Meaning in method: The rhetoric of quantitative and qualitative research. *Educational Researcher, 16*(7), 16-21.

Fischer, C. (1982). Science and the environment of young girls. In J. Ferguson (Ed.), *Who turns the wheel?* Ottawa: Science Council of Canada.

Haggerty, S. (1987). Gender and science achievement: A case study. *International Journal of Science Education, 9*(3), 271-279.
Harding, J. (Ed.). (1986). *GASAT III: Contributions from the third conference of Girls and Science and Technology*. London: Kings College, University of London.
Howe, K.R. (1988). Against the quantitative-qualitative incompatibility thesis or dogmas die hard. *Educational Researcher, 17*(8), 10-16.
Johnson, S., & Murphy, P. (1984). The underachievement of girls in physics: Toward explanations. *European Journal of Science Education, 6*(4), 399-409.
Kahle, J. (Ed.). (1985). *Women in science: A report from the field*. London: Falmer.
Kahle, J., & Lakes, M. (1983). The myth of equality in science classrooms. *Journal of Research in Science Teaching*, 20(2), 131-140.
Keller, E.F. (1985). *Reflections on gender and science*. New Haven: Yale University Press.
Kelly, A. (Ed.). (1981). *The missing half: Girls and science education*. Manchester, UK: Manchester University Press.
Kelly, A. (Ed.). (1987a). *Science for girls?* Milton Keynes, UK: Open University Press.
Kelly, A. (1987b). Does that train-set matter? Scientific hobbies and science achievement and choice. In J. Daniels & J. Kahle (Eds.), *GASAT IV* (Vol. 1). Ann Arbor, MI: University of Michigan.
Kimball, M. (1982). Sex differences in intellectual ability. In J. Ferguson (Ed.), *Who turns the Wheel?* Ottawa: Science Council of Canada.
Klainin, S., & Fensham, P. (1987). Learning achievements in upper secondary school chemistry in Thailand: Some remarkable sex reversals. *International Journal of Science Education, 9*, 217-227.
Klainin, S., Fensham, P., & West, L. (1989a). Successful achievements by girls in physics learning. *International Journal of Science Education, 11*, 101-112.
Klainin, S., Fensham, P., & West, L. (1989b). The superior achievement of girls in chemistry and physics in upper secondary schools in Thailand. *Research in Science & Technological Education, 7*(1), 5-14.
Koshland, D. (1988). Women in science. *Science, 239*, 1473.
Lewis, S., & Davies, A. (1988). *Gender equity in maths and science*. GAMAST Professional Development Manual. Canberra: Curriculum Development Centre.
Lie, S., & Bryhni, E. (1983). Girls and physics: Attitudes, experiences and underachievement. In S. Sjoberg (Ed.), *GASAT II*, Oslo: Science Education Center, University of Oslo.
Linn, M., & Hyde, J. (1989). Gender, mathematics and science. *Educational Researcher, 18*(8), 17-27.
Maccoby, E., & Jacklin, C. (1974). *The psychology of sex differences*. Stanford, Cal.: Stanford University Press.
Manthorpe, C. (1982). Men's science, women's science or science: Some issues related to the study of girls' science education. *Studies in Science Education, 9*, 65-80.
Martin, J.R. (1988). Science in a different style. *American Philosophical Quarterly, 25*, 129-140.
Martin, J.R. (1989). What should educators do about gender bias in science? In K. Tobin & M. Matthews (Eds.), *Proceedings of International Symposium on History and Philosophy of Science and Science Teaching*. Tallahassee, FL: Florida State University.
Marton, F. (1988). Phenomenography: Exploring different conceptions of reality. In D.M. Fetterman (Ed.), *Qualitative approaches to evaluation in education: The silent scientific revolution*. New York: Praeger.
McClintock Collective. (1988). *Getting into gear*. Melbourne: Ministry of Education.
McLarty, J., Noble, C., & Huntley, R. (1988, April). *Effects of item wording on sex bias*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans.
Novak, J. (Ed.). (1987). *Proceedings of the second international seminar: Misconceptions and educational strategies in science and mathematics*. Ithaca, NY: Cornell University.
Parsons-Chatman, S. (1987). Females and physical science: Is tinkering an issue? In J. Daniels & J. Kahle (Eds.), *GASAT IV* (Vol. 1). Ann Arbor, MI: University of Michigan.
Raat, J. (Ed.). (1981). *GASAT I: Contributions from the first conference of Girls and Science and Technology*. Eindhoven: Eindhoven University Press.
Sherman, J. (1978). *Sex-related cognitive differences: An essay on theory and evidence*. Springfield, IL: Thomas.

Sjoberg, S. (Ed.). (1983). *GASAT II: Contributions from the second conference of Girls and Science and Technology*. Oslo: Centre for Science Education, University of Oslo.

Sjoberg, S. (1989). Gender and the image of science. In P. Adey (Ed.), *Adolescent development and school science*. London: Falmer.

Smail, B. (1987). Organizing the curriculum to fit girls' interests. In A. Kelly (Ed.), *Science for girls?* Milton Keynes, UK: Open University Press.

Smith, J. K., & Heshusius, L. (1986). Closing down the conversation: The end of the quantitative-qualitative debate among educational inquirers. *Educational Researcher, 15(*1), 4-12.

Soltis, J. (1984). On the nature of educational research. *Educational Researcher, 13*(10), 5-10.

Taylor, H., Hunt, R., Sheppy, J., & Stronck, D. (1982). *The 1982 British Columbia science assessment: General report*. Victoria, BC: British Columbia Ministry of Education.

Vegting, P. (1987). Concept development in the natural sciences: Stimulating for girls. In J. Daniels & J. Kahle (Eds.), *GASAT IV* (Vol. 3). Ann Arbor, MI: University of Michigan.

von Glasersfeld, E. (1989). Cognition, construction of knowledge and teaching. *Synthese, 80*(1), 121-140.

Wedman, I., & Stage, C. (1983). The significance of contents for sex differences in test results. *Scandinavian Journal of Educational Research, 27*, 49-71.

White, R. (1988). *Learning science*. Oxford: Blackwell.

Education and Society is an established international journal which publishes original research and contributions dealing with major up-to-date issues concerning the relationship between education and society. It provides a new forum, where teachers and scholars throughout the world, working in education and related areas, are able to examine and evaluate current issues and problems in education and society — from a broad and balanced social, cultural and economic perspective. The journal, by focusing on both education and society, hopes to appeal to many educators and scholars from diverse educational backgrounds. Each issue of the journal contains a special symposium concerned with some of the more immediate problems affecting education, educational systems and society in a changing world.

The journal should be of interest to teachers, academics, research scholars, educational administrators and consultants, students in teacher training institutes and social workers.

Contributions are invited (normally no more than 5000 words, typed double spaced in two copies) and should be sent to the Editor, *Education and Society,* James Nicholas Publishers, P.O. Box 244, Albert Park, Australia, 3206.

In order to ensure that articles are of the highest quality, all contributions are submitted to at least two independent referees before acceptance for publication.

The journal appears twice a year, in June and December. These two issues constitute one volume. An annual index is included in the December issue.
1990, Volume 8, **ISSN 0726-2655.**

Subscription rates (payable in advance): one yearUS$90.00. Individuals and elementary and secondary schools US$48.00 (for accelerated surface postage and packing add US$5.00 per issue). Orders and remittances should be addressed to:

James Nicholas Publishers, P.O. Box 244, Albert Park, Australia, 3206.

EDITOR **Dr Joseph Zajda**

CONSULTING EDITORS
Professor Orlando Albornoz, *University of Venezuela*
Professor John Eggleston, *University of Warwick*
Professor Nigel Grant, *University of Glasgow*
Professor Henry Levin, *Stanford University*
Professor David Nyberg, *State University of New York (Buffalo)*
Professor Douglas Sloan, *Columbia University*
Professor Seth Spaulding, *International Bureau of Education, Geneva*
Dr. Joan Whitehead, *University of Cambridge*
Dr. John Wilson, *University of Oxford*
Professor Charles Willie, *Harvard University*

RECENT AND FORTHCOMING ISSUES
Ivan Illich **ON EDUCATION**
Martin Carnoy **EDUCATION AND THE THEORIES OF THE STATE**
Edmund King **CHANGING EDUCATION FOR A CHANGING SOCIETY**
John Eggleston **SCHOOLING AND WORK**
John Martin Rich **MERITOCRACY AND EQUAL EDUCATIONAL OPPORTUNITY**
James Bowen **SCIENCE, EDUCATION AND THE ENVIRONMENT**
Francis Bregha & Don Dawson **WORK, EMPLOYMENT AND LEISURE**
Stanley Aronowitz **WORKING CLASS IDENTITY AND CELLULOID FANTASIES**
Michael O'Loughlin & Malcolm Campbell **TEACHER PREPARATION, EMPOWERMENT AND REFLECTIVE INQUIRY**
John Smyth **TEACHERS-AS-INTELLECTUALS**
Joseph Zajda **EDUCATIONAL REFORMS AND IDEOLOGY IN SOVIET EDUCATION**
H. Svi Shapiro **EDUCATION AND THE CRISIS ON SOCIETY**
Nigel Grant **IMPACT OF COMPARATIVE EDUCATION ON SOCIETY**
Michael Grady & Charles Willie **COMMUNITY PARTICIPATION IN EDUCATION**
Gunnar Berg **THE SCHOOL ORGANIZATION**
Joseph Zajda **CURRICULUM REFORMS IN SOVIET SECONDARY SCHOOLS**

EDUCATION AND SOCIETY

ORDER FORM

☐ **Please enter our subscription to *Education and Society***

We enclose $ .....

☐ **Please enter my subscription under the terms of the Personal Subscription Plan**
(please be sure to give your private address)

We enclose US$ .....

☐ **Please charge my**
Access/American Express/Barclaycard/Diners Club/Mastercard/Visa account number: .....

Card expiry date ..... **Signature** .....

Name .....

Address .....

.....

Signed ..... **Date** .....

The Alberta Journal of Educational Research *Vol. XXXVII, No. 3, September 1991, 241-257*

BIKKAR S. RANDHAWA
University of Saskatchewan

Gender Differences in Academic Achievement: A Closer Look at Mathematics[1,2]

The present study addressed the issue of change in gender differences in academic achievement, more specifically in mathematics, using the effect size as a standard unit of comparison. The effect size was used both within and between years to discern gender differences on specific tests in mathematics and language areas administered to a provincial stratified cluster (classroom) sample of students at grades 4, 7, and 10. Additionally, mathematics and reading data collected in each of three consecutive years from a sample of Grade 10 students were analyzed to determine whether gender differences were consistent and systematic over three years. Results suggest that in grade 4, females increased their achievement advantage on Capitalization and Mathematics Concepts subtests between 1978 and 1985. The female achievement advantage observed in grade 7 in 1978 was significantly reduced by 1985 in Vocabulary, Capitalization, and Mathematics Concepts subtests. In grade 10, males outperformed females in mathematics while females achieved higher than males in language/reading. These grade 10 male/female advantages did not significantly change between 1978 and 1985. Furthermore, in mathematics components, the grade 10 results from one school were consistent and systematic, indicating consistent male advantage in computation, concepts, and problem solving. These differences cannot be attributed to differences in perceived classroom behaviors of males and females by their teachers nor to differences in cognitive structures.

Gender differences in cognitive tasks have intrigued researchers since the turn of this century (Jacklin, 1989). However, since the publication of the Benbow and Stanley (1980) study on mathematically precocious youth, there has been considerable controversy regarding both the extent and the reasons for the gender differences in achievement, more specifically in mathematics involving measurement using standardized tests (Hyde, Fennema, & Lamon, 1990; Randhawa & Hunt, 1987) and using teacher assigned grades (Kimball, 1989). The focus of the present article is on gender differences in mathematics, but it also deals with differences in language related subtests.

Gender Differences

Mathematics

Generally, until about grade 4 either no differences between males and females are observed in mathematics achievement (Callahan & Clements, 1984; Dossey, Mullis, Lindquist, & Chambers, 1988; McKay, 1979; Randhawa & Hunt, 1987) or differences that are observed favor females (Brandon, Newton, & Hammond, 1985; Hawn, Ellet, & Des Jardines, 1981;

Bikkar Randhawa is a professor of educational psychology with current research interests in cognitive psychology and gender-related differences in mathematics learning and instruction.

Randhawa & Hunt, 1987). An exception to these findings was reported by Lewis and Hoover (1986). They observed that on the standardization of the Iowa Tests of Basic Skills, grade 2 boys achieved at a higher level than girls on the problem solving subtest but that this difference decreased in grade 5 and disappeared by grade 8.

Mixed results have been reported for the junior high school grades (years). Either small gender differences in favor of boys (Hilton & Berglund, 1974; Raphael, Wahlstrom, & McLean, 1984; Tsai & Walberg, 1979) or no differences between males and females (Cicirelli, 1967; Connor & Serbin, 1985; Raphael et al., 1984; Randhawa & Hunt, 1987) have been reported. For example, the results of the Ontario portion of the Second International Mathematics Study (Raphael et al., 1984) involving grade 8 students indicated that boys scored significantly higher than girls on 19 items, girls scored significantly higher than boys on 10 items, and there were no differences on 11 items. Moreover, it was found that boys tended to perform better on the more difficult questions. Exceptions to these mixed results are studies of "gifted" youth in which consistent reliable differences in favor of boys were obtained (Benbow & Stanley, 1982; Weiner, 1984). However, it must be kept in mind that in these latter studies gender differences were examined in terms of total SAT scores rather than in terms of item (Raphael et al., 1984) or component score (Randhawa & Hunt, 1987) differences.

On overall math achievement scores throughout the high school years, differences favoring boys are common (Jones, Burton, & Davenport, 1984; Ramist & Arbeiter, 1986; Randhawa, 1988; Randhawa & Hunt, 1987). However, gender differences favoring males are more commonly and consistently found on problem solving components and applications (Carpenter, Lindquist, Matthews, & Silver, 1983; Randhawa, 1988; Randhawa & Hunt, 1987; Swafford, 1980).

Verbal and Language Arts

Gender differences in mathematics achievement favoring males, especially at the high school level and beyond, have received considerable attention and debate (Wittig & Petersen, 1979). However, the consistent female advantage in verbal tests and skills generally throughout the school years (Alberta Department of Education, 1989; Green, 1987; Myerberg, 1986; Randhawa & Hunt, 1987) has not been the source of much debate and concern regarding the causative factors explaining female superiority. Even beyond the high school, consistent but declining female advantage in verbal tests is reported (Wilder & Powell, 1989).

Wilder and Powell (1989) provide an extensive review of the significant research on gender differences in test performance, including studies documenting these differences and theories put forward for the factors related to differential performance. This review includes studies involving undergraduate admission tests, graduate and professional school admission tests, validity studies, national studies, verbal ability tests, quantitative ability tests, and voluntary testing programs. Explanations of test performance differences reported in the literature, according to this survey, include biological, social, psychological, demographic, and educational factors as well

as test characteristics. Some of these explanations are discussed in more detail in the next section in relation to gender achievement differences in mathematics.

Wilder and Powell (1989) conclude that there is a convergence in the literature on gender differences in that males have caught up with females on tests of verbal ability and achievement, while females have narrowed the gap between themselves and their male counterparts on tests of mathematics achievement. Whether similar trends emerge in a provincewide data on language/verbal tests and mathematics tests is one of the objectives of this article. Also, this article explores the underlying causes, albeit empirically and speculatively, of gender differences in achievement in both of these areas.

Explanations of Gender Differences

Psychologists have attempted to explain gender differences in mathematics achievement on the basis of underlying ability (Benbow & Stanley, 1980; 1981; 1982; 1983) and differential course taking (Fennema & Sherman, 1977; Linn & Hyde, 1989). Although Benbow and Stanley speculate about biological causes for their observed gender differences, they had no data to substantiate their conclusion. Also, their studies did not take into account their subjects' differential math-related experiences and assumed that enrollment in equal numbers of mathematics classes would account for this factor (Eccles & Jacobs, 1986; Lips, 1988).

Stylistic information processing, as an alternative explanation to gender differences, was proposed by Dinnel and Glover (1987). Assumed in this explanation are the well-known concepts of storage and retrieval capacity and strategy, knowledge representation, planning, and execution. On the other hand, Fennema and Peterson (1985) suggest that gender-related differences in mathematics might be explained on the basis of the autonomous learning behavior of boys and girls. Choosing to do high-level tasks, working independently on these tasks, persisting in them, and attaining success on them are examples of autonomous learning behaviors. Fennema and Peterson suggest that males have more opportunities than females to engage in these behaviors both in the school and outside. Accordingly, the nature and type of contact between the teacher and the student determine the opportunity for the student for autonomous behaviors. They report that some types of contact enhance autonomous behaviors, but females are not as likely as males to have those contacts. In the same vein, Eccles and Jacobs (1986) attribute social forces as possible determiners of mathematics performance and attitudes.

On the basis of a meta-analysis of 98 studies of gender differences in mathematical tasks reported between 1975 and 1988, Friedman (1989) conclude that "we cannot say with 95% certainty that a gender difference exists in the general United States population of school-age youth" (p. 204). These studies included first graders to twelfth graders. However, a subset of 36 of these studies, in which the performance of high school-age students was examined, indicated that the gender difference in favor of males is diminishing as a function of the year of study (Friedman, 1989). She proposes an

environmental explanation for the decreasing gender differences in mathematics.

Huston and Carpenter (1985) speculate that general gender differences may in fact be present by the time children enter elementary schools. They found from their research with preschool children that girls spend more time in highly structured activities than boys. Structure in this case was determined by (a) the amount of feedback given by adults, and (b) the extent to which the child could observe adults performing activities similar to those in which the child was involved. Huston and Carpenter found that preschool children who were engaged in highly structured activities were more likely to seek approval from adults than those engaged in less structured activities. Therefore, it is probable that gender differences in responsiveness to adults that appear during the preschool years continue into the later years. Hence this tendency of differential responsiveness of males and females may stimulate teachers to react differently to males than to female students.

This article integrates the results from three research studies of gender differences in language and mathematics. The first two, in 1978 and 1985, were based on two provincewide assessments conducted at grades 4, 7, and 10. The third study, in 1987-1989, was based on a three-year assessment of reading and mathematics of grade 10 cohorts in a high school. More specifically, the questions addressed were:

1. What are the patterns of change in the effect sizes for various grades and tests between test occasions? Has the gap between the achievement of males and females in language and mathematics narrowed between 1978 and 1985 for grades 4 and 7 and between 1978, 1985, and 1987-1989 for grade 10?
2. For the grade 10 subjects tested in 1987, 1988, and 1989, are the gender differences in reading and mathematics consistent?
3. Are the classroom behaviors of males and females rated similarly by the students' teachers?
4. How do the factor structures of males and females on reading and mathematics microcomponents compare?

Method

Samples

1978 and 1985 studies. Similar sampling procedures were used for both the 1978 and 1985 studies. A single-stage stratified (by school district) sampling design was used to sample classes from each school jurisdiction in Saskatchewan so that 10% of the classes during the academic years 1978-1979 and 1985-1986 were included. Jurisdictions with fewer than six classes at each of the selected grade levels, 4, 7, and 10, were placed in a separate stratum and sampled at the 10% rate at each grade level. In all cases, all students in each sample class were tested. However, data for a few classes (between 1% and 2% of the classes) were not available due to failure to conduct the testing according to instructions or lost in the mail.

1987-1989 study. A high school in a major city in Saskatchewan was the source of data for a three-year study of reading and mathematics achievement. All those grade 10 students in each of the three years who were present

during the two days of test administration in mid-May were tested. This school serves a fairly large enrollment area and students from all socioeconomic backgrounds attend it. The distribution of students by gender and grade for each study is summarized in Table 1.

Tests

1978 study. Students in grades 4 and 7 in the 1978 study were administered the Canadian Tests of Basic Skills (CTBS), Form 4M (King, 1977). The grade 10 students were administered the Sequential Tests of Educational Progress (STEP), Series 2, Form 2A (Educational Testing Service, 1971). The specific subtests of the CTBS and STEP are listed in Table 2 in the order in which they were administered.

1985 study. The tests used for the 1985 study were the appropriate levels of the CTBS, Form 5M (King, 1982) for grades 4,7, and 10. The specific subtests of the CTBS for each grade and the order of administration are given in Table 2.

1987-1989 study. As shown in Table 2, Reading and Mathematics subtests of the CTBS, Form 5 were administered in the 1987-1989 study.

As well, teachers of Algebra 10 (a grade 10 course) were asked to rate selected students from their classes using the Florida Key (FK) (Purkey, Cage, & Graves, 1973) at the end of school year in June 1989. Altogether, 86 completed FKs were returned. Of these, 43 were for males and 43 for females.

The FK scale contains four subscales: Relating (5 items), Asserting (7 items), Investing (2 items), and Coping (4 items). Each item is rated using a 4-point Likert scale. As a research instrument, the reliability and validity data on this instrument are modest at best. Based on the nominations of 27 teachers of five students who the teachers thought felt "best about themselves as learners" and another five students who the teachers thought "felt badly about themselves as learners," point-biserial correlation coefficients for items with subscales ranged from .57 to .71. An interreliability of .84 for

Table 1
Distribution of Students by Sex, Grade, and Year of Study[a]

Year	Grade					
	4		7		10	
	M	F	M	F	M	F
1978	675	654	790	706	859	900
1985	915	832	837	849	834	753
1987					58	64
1988					58	72
1989					82	79

[a]The number of males and of females included in the 1978 and 1985 studies represented between 9.7 to 10.6% of the provincial enrollment by gender on September 30 in the respective grades.

Table 2
Standardized Tests Administered to the Various Samples and the Order of Administration of the Subtests

Grades 4 and 7		Grade 10		
1978	1985	1978	1985	1987-1989
CTBS[1], Form 4M	CTBS, Form 5M	STEP[2], Form 2A	CTBS, Form 5M	CTBS, Form 5M
Vocabulary		Reading	Reading	Reading
Reading		Mechanics of Writing	Mathematics	Mathematics
Language:			Written Expression	
Spelling		English Expression		
Capitalization			Using Sources	
Punctuation		*Mathematics:*	of Information	
Usage		Basic Concepts		
Mathematics:		Computation		
Concepts				
Problem Solving				
Computation[a]				

[1]CTBS is the Canadian Tests of Basic Skills.
[2]STEP is the Sequential Tests of Educational Progress.
[a]This subtest was included in Form 5M as a separate entity. Priorly, items on computation were included in the Problem Solving subtest.

the composite scale was reported using the data of three teachers who had rated the same 11 students.

Procedure

The testing for the 1978 and 1985 studies was completed during October and November. The tests were administered by the classroom teachers according to a schedule provided by the researcher (see Table 2). For the 1987-1989 studies the tests were administered in mid-May of each year. As shown in Table 2, the Reading test was administered first followed by the Mathematics test in each of the three years.

In June, 1989 the teachers provided data for the selected students on the FK.

Analyses

To answer the first question, the basic statistic calculated was the effect size. This is based on Hyde (1981) and corrected by the factor shown by Hedges (1981) to produce an unbiased estimator of the population effect size for both large and small samples. The corrected effect size for a test is given by

$$d = c\,(Y_F - Y_M)/s,$$

where d is the corrected effect size for a test in a given year, c is the correction factor, Y_M and Y_F are the means of males and females, respectively, and s is

the pooled standard deviation of both groups on the same test. The correction factor c is defined as

$$c = 1 - \frac{3}{4n-9},$$

where n is the total number of males and females. The estimated variance of d is found by

$$s^2(d) = \frac{n}{n_F n_M} + \frac{d^2}{2n}$$ (Hedges & Olkin, 1985, p. 86).

The 1978 and 1985 studies involved fairly large samples of over 1,300 subjects and the combined sample size of the 1987-1989 study was over 400 (see Table 1). With samples as large as these, the variance of the estimate of effect size becomes quite small and hence the estimate is very precise (Friedman, 1989). Therefore, small changes in effect sizes would indicate significant shifts in gender differences in the tested achievement. Furthermore, regardless of the effect size, the estimate of the standard error of an effect size is influenced markedly by the sample size of each group. Hence, as will be noted in the results, the standard errors of effect sizes for various subtests for the same two groups are almost identical.

For the second question, two separate 2 x 3 (gender-by-year) fixed effects multivariate analysis of variance (MANOVA) were performed. In the first, the dependent variables were the CTBS Reading and Mathematics total scores. In the second, the three mathematics component scores—computation (9 items), concepts (20 items), and problem solving (19 items) were the dependent variables. Not normally reported, derivation of these component subscores was based on the classification system provided in the CTBS manual. A similar MANOVA involving the components of the Reading test was not performed because, unlike the Mathematics test, none of the univariate effects for it were significant.

Given nonorthogonality due to unequal sample sizes, the analyses were conducted so that the effects other than the one of interest are first partialled out. This yields a more conservative estimate of the *F*-ratio than if this partialling had not been done.

In order to answer the third question, a MANOVA was performed on the four FK subscales following the same analysis approach as that used for the mathematics components.

The fourth question was addressed by the analyses described below. The CTBS, Level 16, Mathematics and Reading items were grouped according to the process and content classifications as described in the test manual. On each of these 21 microcomponents the score of each subject was computed. The data of male and female students from the high school, combined over three years, were processed separately. First a principal components factor analysis was performed. In each case two principal components were retained. This decision was based on the Cattell's (1966) scree test for the number of factors. The obtained principal factor structures were then subjected to an orthogonal Varimax rotation (Kaiser, 1958). Oblique rotation might be theoretically interesting for the achievement microcomponents, but for the purpose of the present study orthogonal rotation served the function

well. Furthermore, examination of the factor structures from both oblique and orthogonal rotations indicated that there was little difference in the simple structures or the factor interpretations. Lastly, these factor matrices were rotated to a procrustes congruence criterion (Schonemann, 1966) in order to determine whether the factor structures of males and females were invariant (congruent). The females' factor matrix was used as the target for this rotation.

Results

Question 1: What are the patterns of change in the effect sizes for various grades and tests between test occasions? Has the gap between the achievement of males and females in language and mathematics narrowed between 1978 and 1985 for grades 4 and 7 and between 1978, 1985, and 1987-1989 for grade 10?

Grade 4. Effect sizes for gender differences on the CTBS tests for the 1978 and 1985 samples of grade 4 students are given in Table 3. In 1978 females had higher observed means than males on all subtests except the Mathematics Concepts and Problems subtests ($p < .05$). Likewise, except for the Mathematics Concepts, grade 4 female students in 1985 outperformed the grade 4 males ($p < .05$).

The estimation of effect sizes is used in quantitative research syntheses of studies. However, in the present study, these estimated effect sizes are used to determine whether in the time span between the studies the female/male advantage has increased or decreased. Given the standard errors for various subtests in Table 3 for 1978 and 1985, the standard error of the difference between effect sizes for any subtest is .073. This is because the standard error

Table 3
Within Group Standard Deviations, Effect Sizes, and the Standard Deviations of the Effects for Grade 4 Samples

Subtest	SW		Female-Male Effect Size (d)				[d]1985-1978
	1978	1985	1978		1985		
1. Vocabulary	6.316	7.267	.142*	(.055)[a]	.156*	(.048)	.014[b]
2. Reading	11.755	9.081	.103*	(.055)	.210*	(.048)	.107
3. Spelling	7.492	6.900	.355*	(.055)	.426*	(.048)	.071
4. Capitalization	6.613	5.444	.277*	(.055)	.436*	(.048)	.159*
5. Punctuation	7.010	5.494	.335*	(.055)	.351*	(.048)	.016
6. Usage	6.045	5.766	.251*	(.055)	.387*	(.048)	.136
7. Math Concepts	6.135	5.041	−.077	(.055)	.079	(.048)	.156*
8. Math Problems	4.998	5.103	.024	(.055)	.123*	(.048)	.099
9. Math Computation[c]	-	7.858	-	-	.285*	(.048)	-

*$p < .05$.
[a]The value in parentheses is the standard error of the effect size.
[b] 95% confidence interval for the difference is +.143.
[c]This subtest was included in Form 5M as a separate entity. Priorly, items on computation were included in the Problem Solving subtest.

of the effect size and the sample sizes for each subtest for a testing occasion were equal. Therefore, a change in the effect size of +.143 would lead us to conclude that the gap has either increased or decreased ($p < .05$). Using this criterion, it is concluded that the grade 4 average female student has gained further ground over the average male student on the Capitalization subtest between the two testing occasions. Similarly, on the Mathematics Concepts subtest the change between 1978 and 1985 was significant, but it might be attributed to chance because the difference between males and females on each testing occasion was not significant.

Grade 7. Effect sizes for gender differences on the CTBS subtests for the 1978 and 1985 studies are given in Table 4. In 1978, on all except the two mathematics subtests females' means were significantly higher than those of males ($p < .05$).

In 1985, positive and significant effect sizes were obtained on all subtests except Vocabulary, Mathematics Concepts, and Mathematics Problems ($p < .05$). In contrast, on Vocabulary the effect size was negative and significant, suggesting that on this subtest the males did better than the females. As with the grade 4s, the effect sizes for Mathematics Concepts and Problems were not significant. Interestingly, on the Mathematics Computation subtest, which was a part of the Problems subtest in the CTBS battery used in 1978, females outperformed males. Based on the criterion shift of + .14 ($p < .05$) in the effect size, it is evident that females lost ground to males on Vocabulary and Capitalization subtests. Similar to the results of the grade 4 studies on the Mathematics Concepts subtest, the effect size for each testing occasion was not significant. But the difference in the effect sizes between

Table 4
Within Group Standard Deviations, Effect Sizes, and the Standard Deviations of the Effects for Grade 7 Samples

Subtest	Sw		Female-Male Effect Size (d)		$d_{1985-1978}$
	1978	1985	1978	1985	
1. Vocabulary	8.071	8.136	.129* (.052)[a]	−.114* (.049)	−.243*[b]
2. Reading	13.307	9.800	.255* (.052)	.120* (.049)	−.135
3. Spelling	8.990	8.590	.489* (.053)	.471* (.049)	−.018
4. Capitalization	7.208	5.411	.497* (.053)	.340* (.049)	−.157*
5. Punctuation	7.327	5.994	.495* (.053)	.388* (.049)	−.107
6. Usage	5.878	5.264	.416* (.052)	.315* (.049)	−.101
7. Math Concepts	7.940	7.349	.088 (.052)	−.060 (.049)	−.148*
8. Math Problems	4.838	6.017	.012 (.052)	.025 (.049)	.013
9. Math Computation[c]	-	7.411	- -	.229* (.049)	-

$p < .05$.
[a]The value in parentheses is the standard error of the effect size.
[b] 95% confidence interval for the difference is +.143.
[c]This subtest was included in Form 5M as a separate entity. Priorly, items on computation were included in the Problem Solving subtest.

the two occasions on this subtest was significant, which might suggest that females lost ground to males between 1978 and 1985. It might be possible that this result is a chance artifact.

Grade 10. Effect sizes for gender differences on the STEP and CTBS subtests are provided in Table 5 for the three grade 10 samples. Females outperformed males on all four language-related subtests of the STEP ($p<.05$). On the other hand, males outperformed females on the Math Concepts subtest ($p < .05$). On the Math Computation subtest the difference was not statistically significant.

On the CTBS subtests in 1985 all four effect sizes were significantly different from zero ($p < .05$). Moreover, on the Reading, Written Expression, and Using Sources of Information (USI) subtests, females outperformed males, while on the Mathematics subtest males outperformed females. However, for the 1987-1989 sample, the effect size for the Reading subtest was not significantly different from zero. But a significant effect size on the Mathematics subtest with a negative sign, as was the case for the 1985 sample, was observed. Hence the average male outperformed the average female on this subtest.

To compare the results of 1978 and 1985, the average effect sizes across the mathematics subtests and across the language subtests within each year

Table 5
Within Group Standard Deviations, Effect Sizes, and the Standard Deviations of the Effects for Grade 10 Samples

Subtest	SW			Female-Male Effect Size (d)				
	1978	1985	1987-1989	1978	1985		1987-1989	
STEP								
Reading	8.52	-	-	.136* (.048)	-		-	
Spelling	7.70	-	-	.440* (.048)	-		-	
Capitalization	7.00	-	-	.493* (.048)	-		-	
English Expression	8.95	-	-	.331* (.048)	-		-	
Math Computation	9.68	-	-	.032 (.048)	-		-	
Math Concepts	6.30	-	-	−.175* (.048)	-		-	
CTBS								
Reading	-	9.28	9.45	-	.288*	(.051)[a]	.051	(.099)
Mathematics	-	7.98	8.48	-	−.116*	(.050)	−.330*	(.099)
Written Expression	-	8.74	-	-	.546*	(.051)		
Using Sources of Information	-	10.12	-	-	.259*	(.050)	-	
d_{math}				−.071	−.116	−.330		
$d_{language}$				.350	.364	.051		

$p < .05$.
[a] Standard error of the effect size.

were computed and compared. In the case of mathematics, the effect sizes were, respectively, –.071 and –.116. Since the difference between the two was less than .1 in absolute terms, it is concluded that the relative standing of males and females in mathematics did not change. The same conclusion is reached when we compare the average effect sizes of language. The average effect sizes for 1978 and 1985 on the language subtests were, respectively, .350 and .364. Moreover, if the USI subtest, which includes a number of mathematics items, is excluded, the average effect size for 1985 in language-related subtests is .417. The above conclusion is still valid.

Now, if the 1987-1989 results from the single school included in the 1987-1989 study are compared with the average effect sizes computed for the 1978 and 1985 samples, it might be concluded that, at the grade 10 level, females lost ground to males in both language and mathematics. However, such a comparison is not warranted given the differences in sample selection noted earlier. Thus it appears that there is no discernible shift in the achievement patterns of a provincially representative sample of males and females in mathematics and language-related areas between 1978 and 1985.

Question 2: For the grade 10 subjects tested in 1987, 1988, and 1989, are the gender differences in reading and mathematics consistent?

In the first analysis using Reading and Mathematics as dependent variables, the only significant effect was the Gender main effect ($F(2, 406) = 13.53$, $p<.05$). The corresponding univariate analysis revealed that the Gender effect was significant ($F(1, 407) = 11.69, p < .05$) only on Mathematics. The corresponding means for males and females were 30.20 and 27.39, respectively. Since the first-order interaction was not significant, it is evident that the males' advantage in achievement in mathematics over the females' was consistent and systematic over the three years.

The Gender main effect was also significant in the MANOVA analysis of the three mathematics component scores ($F(3, 405) = 4.64, p<.05$). Furthermore, the univariate Gender main effects (df = 1, 407) on Computation ($F=8.56$, $p<.05$), Concepts ($F=10.95$, $p<.05$), and Problem Solving ($F=12.08$, $p<.05$) were reliable. Again, the first order interaction between Gender and Year was not significant. On each of the three mathematics components, males' mean achievement was higher than that of females' in a consistent and systematic fashion over the three years considered (see Table 6).

Question 3: Are the classroom behaviors of males and females rated similarly by their teachers?

Means and standard deviations of males and females for the subscales of FK are given in Table 7. The MANOVA results indicated that the Gender effect was not significant. Hence it appears that the classroom behaviors of males and females were rated similarly by their teachers.

Question 4: How do the factor structures of males and females on reading and mathematics microcomponents compare?

Application of the scree test to each of the principal component solutions revealed two factors for both males and females. The congruence matrices of females and males and the commonalities of the variables are given in Table

Table 6
Means and Standard Deviations on the Mathematics Components of Grade 10 Students

Component	Year 1		Year 2		Year 3	
Males n	58		58		82	
Computation	6.1	(1.9)[a]	6.1	(2.0)	6.0	(2.1)
Concepts	12.1	(4.3)	13.2	(4.2)	12.6	(4.4)
Problem Solving	11.1	(3.8)	12.3	(3.4)	11.4	(4.2)
Females n	64		72		79	
Computation	5.3	(1.6)	5.6	(2.0)	5.7	(1.5)
Concepts	10.4	(3.6)	12.4	(3.8)	11.2	(3.4)
Problem Solving	9.7	(3.1)	10.9	(3.6)	10.4	(3.3)

[a]Standard deviation.

8. As indicated by the Tucker congruence coefficients for each pair of factors, .97 and .98 respectively, the two pairs of factors were highly congruent. This suggests that, in general, the operative common structures of males and females for handling language and mathematics microcomponents are quite similar. But the two principal factors capture 41.6% and 46.8% of the total variance in the set of 21 variables for females and males respectively. Hence there is enough idiosyncratic variance that could produce differences in achievement on these two tests.

The first factor in each case can be interpreted to be the language and the second the mathematics factor. However, the language factor for both females and for males accounts, respectively, for 81.9% and 84.4% of the common variance in the set of variables. Moreover, the two factor structures are not good simple structures because in each set of variables a number load on both factors with a loading greater than .30. Taken the above considerations into account and that of parsimony, it may well be that there is only a single common factor in each set and which could parsimoniously be labeled as the achievement factor. This is not to deny that there may be specific cognitive components which are more instrumental for doing well on certain

Table 7
Means and Standard Deviations of Males and Females on FK Subscales

Subscale	Male (n = 43)		Female (n = 43)	
Relating	20.2	(2.6)[a]	20.6	(3.0)
Asserting	23.4	(7.8)	24.5	(4.5)
Investing	5.6	(2.6)	5.4	(2.2)
Coping	14.4	(3.5)	14.0	(3.9)

[a]Standard deviation.

Table 8
Congruence Factor Matrices of Males and Females

Variable	Females			Males		
	h^2	I	II	h^2	I	II
Textbook Factual Material	69[a]	78	28	66	75	31
Textbook Inferential Material	66	74	32	73	82	22
Textbook General Material	45	67	03	53	72	13
Everyday Factual Material	59	72	26	66	75	32
Everyday Inferential Material	66	72	38	66	76	29
Everyday General Material	58	73	21	52	65	32
Operational Computation	45	32	59	52	36	63
Equivalent Forms and Order Concepts	50	24	67	62	34	71
Equivalent Forms and Order Problem Solving	17	–04	41	39	–17	60
Common Applications Computation	16	02	41	36	08	60
Common Applications Concepts	40	26	57	36	26	54
Common Applications Problem Solving	55	43	61	66	41	70
Algebra Computation	35	44	40	23	14	46
Algebra Concepts	19	05	44	36	31	52
Algebra Problem Solving	29	27	47	18	24	35
Geometry and Measurement Computation	19	37	23	20	27	35
Geometry and Measurement Concepts	28	21	49	43	39	53
Geometry and Measurement Problem Solving	44	11	65	45	20	64
Statistics, Graphs, and Table Concepts	38	42	45	44	38	54
Statistics, Graphs, and Table Problem Solving	26	31	40	40	42	46
Basic Math Problems Concepts	50	38	60	50	40	58
Eigenvalue		7.17	1.58		8.30	1.54
Total Commonality	8.75					
Tucker Coefficient		97	.88			
Common Variance		81.9	18.1		84.3	15.7

[a]Leading decimal places are omitted.

cognitive tasks than on others, but that these various components are united by a common cognitive orientation. It would appear from our evidence that males and females are not differentiated on cognitive orientation.

Discussion and Conclusions

The studies reported here have examined gender-related differences in language and mathematics in grades 4, 7, and 10. An interesting picture emerges. At the grade 4 level in 1978, girls outperformed boys in the language and verbal subtests but there were no differences on the mathematics subtests. However, in 1985, females outperformed males on all except the Mathematics Concepts subtest. Comparison of the gender effect sizes for grade 4 students in 1978 and 1985 indicated that females had increased their ad-

vantage over males on the Capitalization subtest. This finding is not in conformity with the conclusion of Friedman (1989), based on a meta-analysis, and of Wilder and Powell (1989), derived from a review of the literature. They concluded that the general gap favoring males in mathematics achievement is narrowing with the passage of time. Moreover, Wilder and Powell (1989) concluded that males have caught up with females in tests of verbal ability and achievement. This does not appear to be the case with grade 4 students in language and verbal subtests. Additionally, females either outperformed males on the mathematics subtests or there were no differences.

The opposite of this trend and consistent with Wilder and Power's (1989) conclusion was the case for our grade 7 cohort studies. Females were significantly higher than males on all the language-related subtests in 1978 but on the two mathematics subtests there were no significant differences. However, in 1985, females were still significantly higher than males on all the language-related subtests except Vocabulary, on which males were significantly higher than females. Furthermore, females were significantly higher than males on Mathematics Computation subtest in 1985. Again, contrary to Friedman's (1989) conclusion, but in the opposite direction to what was observed in grade 4, females had lost ground in their advantage over males on Vocabulary and Capitalization subtests. Furthermore, on the Mathematics Concepts subtest, even though the effect sizes in 1978 and 1985 were not significant, the difference in the effect sizes was significant. This might suggest that females had lost ground on Mathematics Concepts subtest in 1985. But caution should be exercised with this interpretation because the result might be a chance artifact.

Consistency with Friedman's (1989) and Wilder and Powell's (1989) conclusion of gender differences in achievement was found in grade 10 cohort studies. Males outperformed females in mathematics and females outperformed males on language-related subtests in 1978 and 1985. In the case of the male advantage in mathematics, the results of the grade 10 study, which included students from only one high school, indicated that the male advantage was consistent and systematic during the three-year period considered.

In spite of the narrowing gap in achievement in language-related areas observed in grade 7, the gap in this area still favors females at grade 10. But the reversal in mathematics achievement differences at grade 4, mostly favoring females, has occurred at the grade 10 level. The gap has narrowed at grade 7 but by grade 10, consistent and systematic differences in all three mathematics components, Computation, Concepts, and Problem Solving, were observed. Preliminary results from a study of mathematics achievement at grade 12 and postsecondary levels seems to confirm this systematic and consistent gender difference in favor of males (Randhawa & Beamer, in progress; Randhawa, Beamer, & Ng, in progress).

How do we explain the differences? At the grade 10 level, we attempted to determine if teachers' ratings of males' and females' classroom behaviors would provide a clue. It seems that teachers found coping, investing, asserting, and relating behaviors of males and females quite similar. But caution is in order because the data set was small and the instrument does not appear

to have satisfactory reliability. Also, the attempt to determine if different cognitive structures of males and females contributed to processing differences that in turn might have led to observed differences did not bear differential evidence. The fact that the two common factors turned out to be invariant suggests that cognitive processing structures of males and females are similar. Because the common factors account for only less than half of the variance, an argument can be made that there might be other idiosyncratic mechanisms of males and females that account for systematic achievement differences. These idiosyncratic mechanisms could include the autonomous learning behavior (Fennema & Peterson, 1985) or the social forces and motivations (Eccles & Jacobs, 1986). Further systematic research of classroom processes and individual learner characteristics might shed further light on the sources of the gender differences. In the meantime, instructional processes must provide for equitable access and opportunity for both males and females to profit from them optimally.

Notes

1. This paper was prepared while the author had a research time stipend from the Social Sciences and Humanities Research Council of Canada. This support is gratefully acknowledged. However, the opinions in and substance of the paper are the sole responsibility of the author and do not imply an endorsement by the funding agency.
2. Sheryl Yuzik, Jason Randhawa, and Lisa Randhawa assisted in the preparation of this paper. Their contribution is much appreciated. However, I am fully responsible for any errors and conceptual fallacies and problems that may still exist in this paper. Also, the cooperation of the school administration and teachers is gratefully acknowledged. We are grateful to the students who supplied the data and without whose participation this paper would have been nothing but a dream. Helpful comments of an anonymous reviewer and of Todd Rogers, the guest editor of this section, are gratefully acknowledged. Send requests for reprints to Dr. Bikkar S. Randhawa, Department of Educational Psychology, College of Education, University of Saskatchewan, Saskatoon, Canada S7N OWO.

References

Alberta Department of Education. (1989). *Achievement testing program. Provincial report: June 1989*. Edmonton, AB: Author. (ERIC Document Reproduction Service No. ED 316577)

Benbow, C.P., & Stanley, J.C. (1980). Sex differences in mathematics ability: Fact or artifact? *Science, 210*, 1262-1264.

Benbow, C.P., & Stanley, J.C. (1981). Mathematical ability: Is sex a factor? *Science, 212*, 118, 121.

Benbow, C.P., & Stanley, J.C. (1982). Consequences in high school and college of sex differences in mathematical reasoning ability: A longitudinal perspective. *American Educational Research Journal, 19*, 598-622.

Benbow, C.P., & Stanley, J.C. (1983). Sex differences in mathematical reasoning ability: More facts. *Science, 222*, 1029-1031.

Brandon, P.R., Newton, B.J., & Hammond, O.W. (1985, April). *The superiority of girls over boys in mathematics achievement in Hawaii*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Callahan, L.G., & Clements, D.H. (1984). Sex differences in rote-counting ability on entry to first grade: Some observations. *Journal for Research in Mathematics Education, 19*, 598-622.

Carpenter, T.P., Lindquist, M.M., Matthews, W., & Silver, E.A. (1983). Results of the third NAEP mathematics assessment: Secondary school. *Mathematics Teacher, 76*, 652-659.

Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245-276.

Cicirelli, V.G. (1967). Sibling constellation, creativity, IQ, and academic achievement. *Child Development, 38*, 481-490.

Connor, J.M., & Serbin, L.A. (1985). Visual-spatial skill: Is it important for mathematics? Can it be taught? In S.F. Chipman, L. R. Brush, & D. M. Wilson (Eds.), *Women and mathematics: Balancing the equation* (pp. 151-174). Hillsdale, NJ: Lawrence Erlbaum.

Dinnel, D.L., & Glover, J.A. (1987, April). *Processing differences between competent and less competent mathematics students*. Paper presented at the American Educational Research Association, Washington, DC.

Dossey, J.A., Mullis, I.V.S., Lindquist, M.M., & Chambers, D.L. (1988). *The mathematics report card: Are we measuring up? Trends and achievement based on the 1986 National Assessment.* Princeton, NJ: Educational Testing Service.

Eccles, J.S., & Jacobs, J.E. (1986). Social forces shape math attitudes and performance. *Signs: Journal of Women in Culture and Society, 1*, 367-380.

Educational Testing Service. (1971). *Sequential tests of educational progress (STEP, Series II): Teacher's manual for administering, scoring, and interpreting grades 10-12.* Princeton, NJ: Author.

Fennema, E., & Peterson, P. (1985). Autonomous learning behavior: A possible explanation of gender-related differences in mathematics. In L.C. Wilkinson & C.B. Marrett (Eds.), *Gender influences in classroom interaction* (pp. 17-35). Orlando, FL: Academic Press.

Fennema, E., & Sherman, J. (1977). Sex-related differences in mathematics achievement, spatial visualization, and sociocultural factors. *American Educational Research Journal, 14*, 51-71.

Friedman, L. (1989). Mathematics and the gender gap: A meta-analysis of recent studies on sex differences in mathematical tasks. *Review of Educational Research, 59*(2),185-213.

Green, D.G. (1987, August). *Sex differences in item performance on a standardized achievement test battery*. Paper presented at the meeting of the American Psychological Association, New York. (ERIC Document Reproduction Service No. ED 291806)

Hawn, H.C., Ellet, C.D., & Des Jardines, L. (1981, April). *Differences in mathematics achievement between males and females in grades 1-3*. Paper presented at the annual meeting of the Eastern Educational Research Association, Philadelphia.

Hedges, L.V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107-128.

Hedges, L.V., & Olkin, R. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hilton, T.L., & Berglund, G.W. (1974). Sex differences in mathematics achievement: A longitudinal study. *Journal of Educational Research, 67*, 232-237.

Huston, A.C., & Carpenter, C.J. (1985). Gender differences in preschool classrooms: The effects of sex-typed activity choices. In L.C. Wilkinson & C.B. Marrett (Eds.), *Gender influences in classroom interaction* (pp. 143-165). Orlando, FL: Academic Press.

Hyde, J.S. (1981). How large are cognitive gender differences?: A meta-analysis using w^2 and d. *American Psychologist, 36*, 892-901.

Hyde, J.S., Fennema, E., & Lamon, S.J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139-155.

Jacklin, C.N. (1989). Female and male: Issues of gender. *American Psychologist, 44*, 127-133.

Jones, L.V., Burton, N.W., & Davenport, E.C., Jr. (1984). Monitoring the mathematics achievement of black students. *Journal for Research in Mathematics Education, 15*, 154-164.

Kaiser, H.F. (1958). The Varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23*, 187-200.

Kimball, M.M. (1989). A new perspective on women's math achievement. *Psychological Bulletin, 105*, 198-214.

King, E.M. (1977). *Canadian tests of basic skills, metric edition, levels 9-14, form 4M, teachers guide*. Toronto, ON: Nelson.

King, E.M. (1982). *Canadian tests of basic skills, high school, multilevel edition, levels 15-18, form 5, teachers' guide*. Toronto, ON: Nelson.

Lewis, J.C., & Hoover, H.D. (1986, April). *Differential prediction of academic achievement in elementary and junior high school by sex*. Paper presented at the annual meeting of the American Educational Research Association, Montreal.

Linn, M.C., & Hyde, J.S. (1989). Gender, mathematics, and science. *Educational Researcher, 18*(8), 17-19, 22-27.

Lips, H.M. (1988). *Sex and gender: An introduction*. Mountain View, CA: Mayfield Publishing.

McKay, L.R. (1979). Sex differences in cognitive styles and mathematics achievement in fourth and eighth graders. (Doctoral dissertation, University of Texas at Austin, 1978). *Dissertation Abstracts International, 39*, 4103A.

Myerberg, N.J. (1986). *The annual report of California Achievement Tests results, 1985-1986*. Rockville, MD: Department of Educational Accountability, Montgomery County Public Schools. (ERIC Document Reproduction Service No. ED 286894)

Purkey, W.W., Cage, B.N., & Graves, W. (1973). The Florida key: A scale to infer learner self-concept. *Educational and Psychological Measurement, 33*, 979-984.

Ramist, L., & Arbeiter, S. (1986). *Profiles of college-bound seniors: 1985*. New York: College Entrance Examination Board.

Randhawa, B.S. (1988). Basic skills and macro- and micro-analysis of mathematics achievement of grade ten students as a function of gender and locale. *British Educational Research Journal, 14*(2), 141-148.

Randhawa, B.S., & Beamer, J.E. (in progress). Sex differences in first year university mathematics.

Randhawa, B.S., Beamer, J.E., & Ng, C.F. (in progress). Sex differences in mathematics in relation to contextual variables.

Randhawa, B.S., & Hunt, D. (1987). Sex and rural-urban differences in standardized achievement scores and mathematics subskills. *Canadian Journal of Education, 12*, 137-151.

Raphael, D., Wahlstrom, M., & McLean, L. (1984). Results from the second international mathematics study: Are boys better at math than girls? *Orbit, 15*(2), 6-9.

Schonemann, P.A. (1966). A generalized solution of the orthogonal procrustes problem. *Psychometrika, 31*, 1-10.

Swafford, J.O. (1980). Sex differences in first-year algebra. *Journal for Research In Mathematics Education, 11*, 335-346.

Tsai, S.L., & Walberg, H.J. (1979). Mathematics achievement and attitude productivity in junior high school. *Journal of Educational Research, 76*, 267-272.

Weiner, N.C. (1984). Cognitive aptitudes, personality variables, and gender difference effects on mathematical achievement for mathematically gifted students (Doctoral dissertation, Arizona State University, 1983). *Dissertation Abstracts International, 44, 3621A.*

Wilder, G.Z., & Powell, K. (1989). *Sex differences in test performance*. (College Board Report No. 89-3). New York: College Entrance Examination Board. (ERIC Document Reproduction Service No. ED 308331)

Wittig, M.A., Petersen, A.C. (Eds.). (1979). *Sex related differences in cognitive functioning: Develoemental issues*. New York: Academic Press.

INTERCHANGE

A QUARTERLY REVIEW OF EDUCATION

- a leading international journal on the critical issues facing educators in Canada and abroad
- a recipient of the prestigious CAFE award for the most original work in educational foundations
- four issues per year feature scholarly articles, critical notices, book reviews, and debates
- recent special issues include Creativity, Thought, and Mathematical Proof; History, Science and Science Teaching; and Understanding Literacy (guest edited by David Olson)
- Edited by Ian Winchester

Regular One-Year Rates for 1990: Institutions $50.00; Individuals $35.00. For 1991: Institutions $55.00; Individuals $35.00. To order send cheque, purchase order, or complete MasterCard or VISA information to OISE Publishing, 252 Bloor St. W., Toronto, Ontario, Canada M5S 1V5.

Published by The Ontario Institute for Studies in Education, Toronto, Canada

The Alberta Journal of Educational Research *Vol. XXXVII, No. 3, September 1991, 259-275*

PERSPECTIVES

The Burt Controversy: An Essay Review of Hearnshaw's and Joynson's Biographies of Sir Cyril Burt[1]

JOHN HATTIE
University of Western Australia

When the controversy relating to Sir Cyril Burt broke in the mid 1970s, the academic climate was most conducive to a scandal. Sir Cyril Burt was associated with defenders of the English class system; many were criticizing those who believed heredity was a major contributor to individual differences; the furor had not abated regarding the 1969 Jensen article which included comments about the genetic reasons for black-white IQ discrepancies and had used Burt's work; and the issues of race, genetics, and education were mixed with politics and demonstrations. Many academics were accused of racism and elitism and of concocting data to support the status quo; that is, the hereditarian viewpoint. It is perhaps not surprising that someone who was would be accused of fraud, especially if that someone was disposed toward the heredity view, seen as elitist, of the establishment and traditional school, and preferably dead.

The incidence of fraud is not known, although over 90% of academics in one study cited instances of cheating known to them (Saint John-Roberts, 1976). Broad and Wade (1982) claimed it was far less, but not negligible. Among them Broad and Wade (1982), Gould (1981), Grafton (1990) and Kohn (1986) cite over 100 cases of academic fraud. No matter what the incidence, it takes only a few cases to be publicized for the credibility of science to be damaged. Burt became psychology's biggest test.

The Accusations

The rumors about Sir Cyril Burt, Britain's most famous psychologist, began in the early 1970s. Jensen (1974) published an article questioning some of Burt's data on twins, but the rumors were fed mostly by Kamin in his various talks and subsequently by his book (1974). Kamin's book overstated the case against the heredity viewpoint and included virulent attacks on those who supported it, particularly Burt, although he stopped short of accusing Burt of fraud. He concluded that "the numbers left behind by Professor Burt are simply not worthy of our scientific attention" (p. 47). When a London *Sunday Times* journalist broke the story on the front page on October 24, 1976 with

---

John Hattie is a professor of education. His research interests are in cognition, creativity, psychometrics, and meta-analysis.

the use of the terms "fraud," "false data," and "invented crucial data," the academic community could no longer stay silent.

There were many accusations and counteraccusations that immediately followed, but it is notable that most of these lacked evidence. It was known that Hearnshaw,* an established British historian of education who had begun to write an official biography of Burt, would now include a commentary on the "scandal." Hearnshaw had the necessary competencies to write the biography—a capacity for absorbing facts, a capacity for stating them, and a point of view (Strachey, 1931).

He made six major claims about Burt. First, Burt could not have possibly collected some of his postwar data and thus he fabricated them by inventing data to fit his theories. Hearnshaw (1979) cited McAskie (1978) who claimed that Burt's data was suspiciously close to normal distributions and argued that Burt's "besetting weakness was to rely on statistical manipulation rather than empirical investigation, forgetting that poor data cannot yield sound conclusions" (p. 70). Second, Burt's crude estimates of parental IQ in the 1920s were passed off as more accurate in the 1960s. The suggestion was that the estimates were adjusted to accord with Burt's preconceptions. Hearnshaw concluded that "the table of IQ scores and social class gradings was an elaborately constructed piece of work, and we are forced to the conclusion that he simply did not possess detailed data, at any rate for the whole sample of his separated MZ twins" (p. 247). Third, correlations changed when the sample sizes did not, and correlations remained invariant (to the third decimal point) when the sample sizes changed. Hearnshaw concluded that the accounts "given in Burt's published papers were false; and that a measure of deception was certainly involved" (p. 241). Fourth, some of Burt's co-workers were fictitious. The claim is that some of his co-workers (particularly Howard and Conway) were apparently invented to justify his collecting data, or so that Burt could further promulgate his views, defend some of his priority claims and boost the stature of his work. Fifth, Burt changed other researcher's contributions to the journal he edited to reflect better his views and contributions. Evidence is provided that Burt did this with the contributions of Eysenck, Clarke and Clarke, and Wrigley. Sixth, Burt altered the history of factor analysis to claim priority over some contributors. Hearnshaw documents the case that Burt arrogated to himself the whole of Spearman's fame, became increasingly obsessed with questions of priority, and unquestionably misrepresented Spearman's position. Moreover, Burt gave Pearson pride of place such that "Spearman was dethroned, and Burt himself elevated in his place! It seems almost a classic example of the killing of the father king" (pp. 176-177). "He falsified history in the interest of self-aggrandizement. That he was guilty of malfeasance there can be no reasonable doubt. The only question at issue can be, was he guilty *simpliciter* or guilty with diminished responsibility as a result of pathological influences?" (p. 180). Hearnshaw answered the question by supporting the latter argument.

*Unless otherwise stated, references to Hearnshaw and Joynson refer to Hearnshaw's 1979 book and Joynson's 1989 book.

Hearnshaw concluded that the data Burt used for his calculations were poor and unreliable, that he made a great many unexplained, careless and inconsistent adjustments to the raw scores of his tests with the results that the figures are quite improbable. He ultimately applied sophisticated statistical techniques to scientifically almost worthless data, with disastrous results. In three instances, Hearnshaw finds Burt guilty of deception: He falsified the early history of factor analysis; he produced spurious data on MZ twins; and he fabricated figures on declining levels of scholastic achievement.

Burt's Motives

Hearnshaw traced Burt's pathology from the early "marks of a delusional system, growing from small beginnings into a blind and warping compulsion" (p. 177). Hearnshaw attributed this pathology to seven causes.

First, Burt's onset of Meniere's disease, diagnosed in 1941, tended to incapacitate him. Hearnshaw noted that those with Meniere's disease typically were obsessional compulsive types with marked psychosomatic lability. Second, Burt's competencies were declining with age. Burt complained of "senile paragraphia" (surely this was a joke!) and there are isolated instances of misplaced words in his notes and letters to his sister. As he was often his own editor, Hearnshaw claimed that he was deprived of a very necessary corrective. "Carelessness, then, partly the result of haste and partly the result of emotional involvement and declining powers, was assuredly a contributory ingredient in the final product" (p. 242). Third, there were warnings in his early childhood as he was forced to survive by observing his peers and enemies closely, "to keep his feelings to himself, to bluff it out, and to outmaneuver those who tried to molest him" (p. 273). "In the lost boyhood of Cyril Burt psychology was betrayed" (p. 291). Fourth, he was suffering other setbacks: his marriage, the breakdown of his health, his difficulties following retirement from the chair, loss of control over the statistical journal he helped found and nurture; the gradual erosion of the system of selective secondary education that he had done so much to promote; new modes of thinking and younger rivals were ousting him from center stage; and the doctrines he believed in were being rejected. These setbacks, according to Hearnshaw resulted in his gamin personality (living by treachery and underhand trickery typical of London slum children) coming to the fore and he could appear "quite unscrupulous in the expedients he employed" (p. 276). Fifth, many of Burt's papers were left at University College during the war, and subsequently were bombed in the blitz. Burt subsequently invented the "lost" data. Hearnshaw is certain that all Burt's data were lost (p. 250) but latter he claimed that the greater part was not destroyed (Hearnshaw, 1990). Sixth, Hearnshaw claimed that Burt collected no new data after 1955 as he had neither the physical fitness, the assistants, nor the resources to carry out more research. Data attributed to post-1955 were fabricated. Seventh, Hearnshaw charged that Burt could be slippery and unscrupulous and had an egomaniac demeanor. He would "misrepresent opponents and blur the issues. Beneath a polite exterior and apparent reasonableness was a steely determination to get the better of the argument, and to humiliate his op-

ponents" (p. 206). Moreover, he could be "hostile, cantankerous, and, if need be, unscrupulous" (p. 270).

Overall, Hearnshaw's biography appeared to be a letdown for the great man: a letdown from preeminence to scoundrel, from innovator to the delinquent. Hearnshaw convinced the majority and the reviews were full of praise. It was, claimed Broad and Wade (1982), a sympathetic and subtly drawn portrait, and Gould (1981) called it splendid.

The Counterclaims

Hearnshaw's views were unchallenged for a decade, until Joynson (1989) published his book which concluded that "the evidence which has so far been presented is insufficient to support the accusations which have been made. It does not follow that Burt's behavior was invariably commendable, and that he possessed no weaknesses of character. But the gross misbehavior of which he has been so widely accused has not, in my opinion, been established. A grave injustice has been done" (p. ix).

Joynson was particularly critical of Hearnshaw, as unlike Flaubert's ideal biographer who is "everywhere present and nowhere present," Hearnshaw does not stay in the background. Joynson claimed that Hearnshaw used flimsy evidence, did not locate primary references, has a tenuous grasp of Burt's procedure, omitted words in quotations that lead to a distorted picture, depended on gossip about long past incidents entirely unsupported by adequate evidence, does not know what Burt is talking about, failed to treat evidence critically, used unverifiable speculation about superfluous absurdities, too often accepted hostile comments of Burt's enemies without examination, was selective of evidence, was one-sided, and made innumerable errors. Joynson offered countering evidence for each of the six major charges made by Hearnshaw.

Regarding Burt's guessing of parental and occupational IQ and then passing these off as real, Joynson agreed that Burt often gave less information than was desirable, but contrary to Hearnshaw's claims, Burt did acknowledge the unreliability of his data with comments such as "pilot enquiry," "less thorough and less reliable" and pointed out to correspondents the problems of these data. It seems to me that Burt on more than one occasion noted the unreliability of these data, using terms such as the "comparisons are highly precarious," the figures cited are "the best we can get in the way of objective data," "this caution (the difficulty of procuring samples which can be safely regarded as comparable) applied equally to the figures set out above," and we need to "make liberal allowance for" the way the data were collected. Researchers still estimate parental IQ in ways similar to Burt's procedures (see Jensen, 1974).

Joynson did see some case for criticism of Burt's (1969) description and presentation of material in his "Intelligence and Heredity" article. He explained, however, how Burt could have obtained his data and noted that the figures that Hearnshaw claimed were fabricated are precisely those that have the least tendency to support Burt's argument. "If Burt had been inventing the figures he would surely have chosen numbers which would strengthen his argument rather than weaken it" (p. 209). It seems that Burt

collated means from various publications by himself, his research students (many of whom he acknowledged were now teachers), and perusing manuals of tests, such as Schonell, Vemon, Weschler, or the National Foundation for Educational Research. This crude attempt at meta-analysis shocks Hearnshaw but was a precursor to a revolution of the 1980s (Glass, McGaw & Smith, 1981). Many footnotes and references, not discussed by Hearnshaw or Joynson, outlined where Burt could have gleaned the data. For example, Hearnshaw considered that Burt could not have obtained the 1955 data as "by then Burt had no right of entry to London schools" (p. 257) and concluded that these data were fabricated. Burt detailed (perhaps overdetailed) the collection of such data in 1954 with the help of Mr A.M. Reid, of over 300 London boys at the age of 9.0-10.0, and four years later retested the same group. Further, Audley and Rawles (1990) have claimed that Burt had access to data gathered by the National Foundation for Educational Research from 1955 to 1965 for a survey of ability and attainment. It is claims like these that lead to difficulty in presuming that Burt fabricated his data.

That the correlations suspiciously changed while the sample size did not, or remained invariant when the sample sizes did change, appears to be the hardest claim to overturn, and I confess it took me several readings of this chapter to understand fully Joynson's explanations. He first corrects some anomalies such as typographical errors noted by Burt (Ron Hambleton kindly lent me his copies of Burt's papers which he obtained in response to an advertisement from Burt's secretary shortly after Burt's death. These offprints include corrections and support Joynson's claims. Also see Jensen, 1974, Skanes, 1978). Joynson then argued that the .77 correlation for MZ twins reared apart in 1943 and the .771 in 1955 may be different although the sample size increases from 15 to 21, as the former could be anywhere between .766 and .774; that the .54 for correlation for DZ twins in 1943 and .542 in 1955, though the sample size increased from 157 to 172, could be the result of cumulative addition of new data to the previous data. Thus both are plausible. This is difficult to believe, although not impossible. There is a more obvious conclusion, though it too presents a problem. Burt (1955) cited 21 *pairs*, but Conway (1958) cited 42 *cases*, so the data are probably the same. The problem is that the correlations changed (although the mean change is only .008). In Burt (1966a) it is clear that there are 53 *pairs* (mean change compared to 1955 is .016). It seems believable that Burt could have located an extra 32 pairs in the 11 years (particularly if his acknowledgments to students and colleagues are to be believed, e.g., Burt, 1966a, p. 141).

All anomalies are traced to two papers of which 29 correlations are invariant out of the 64 reported. The invariant correlations, however, "are not to be found concentrated in the crucial areas of tests of intelligence, nor in relation to MZ twins, as might have been expected if Burt was primarily concerned with fabricating evidence about the inheritance of intelligence of twins. Rather, they appear predominantly among the physical measurements, and are scattered across every category of subject" (Joynson, p. 146). Only four concern intelligence, of which two have already been questioned, above. Moreover the coefficients changed in the direction opposite to Burt's hypothesis: "the changes rather suggest a strengthening of the en-

vironmental influence, a strange effect for a scheming hereditarian to produce" (p. 159).

If Burt wanted to stretch his case he would have changed the intelligence coefficients; he would not have been so stupid as to leave such obvious consistencies, would have faked more carefully, and would have changed the correlations in the direction of his hypotheses: "A guilty man would never have made such an error" (Joynson, p. 157). If he had faked them, why did he not point to the remarkable consistencies as evidence in his favor? Why invent numbers and then not reap the rewards? Thus Joynson concluded that there has been no reliable indication of fabrication whatever.

With respect to the existence of Burt's co-workers, Joynson cited Burt's housekeeper that Burt was fulfilling a promise to credit the "missing" co-workers with work done earlier, long before he wrote the articles. Joynson agreed with the critics that Burt frequently wrote letters, comments, and book reviews and published them in his journal under various pseudonyms. He dismissed the claim, however, as an unprofitable basis for sober conclusions but an admirable subject for sensational speculation in the media. It may be possible to attribute some of the sole authored articles by Howard and Conway to Burt (e.g., by computer concordance procedures), although it is probably not worth the trouble. After reading the various contributions by Howard and Conway, what is striking is not the bias and support of Burt's contributions as claimed by Hearnshaw, but the opposite. The papers and reviews are sober, well argued, extensively footnoted; and the style is unmistakably extremely close to Burt's.

Howard has been sighted by some of Burt's peers, by at least one twin who was tested by her, and according to Burt she emigrated before 1950. It is possible that she is the same Howard who was once a member of the British Psychological Society. That no one answered a request in the London *Times* is far from convincing evidence that she or Conway do not exist. It is easy to imagine many people in the back blocks of Britain, Canada, or Australia who have yet to receive their 1976 editions of the *Sunday Times*; and it is possible that Miss Howard and Miss Conway married and changed their last names. As Hearnshaw claimed, they were probably LCC social workers or care committee workers and thus it is not surprising that they would be difficult to locate.

That Burt changed others' contributions to the journal he edited is not challenged; whether these changes aimed to reflect better his view is. It is probable that Burt performed his editorial duties with relish, but most of the documented cases are trivial, misleading, or false. For example, one case documented by Hearnshaw is that of Clarke, who claimed that Burt had altered his and his spouse's work such that they "did not recognize them as the same summaries ... [and] the whole emphasis of 'our' articles was slanted (against Eysenck)" (p. 148). Joynson countered that when compared with the actual theses, the published summaries (not articles) "appear to me to be accurate and well-balanced accounts of the main points contained in the theses" (p. 246), and if changed at all it was changed to be more favorably inclined toward Eysenck than was the thesis. Changing others' work or claiming credit for others' work is a common claim, especially of students

about mentors. This is an often discussed issue that is not unique to Burt. There were also many who were appreciative of the attentions that Burt gave to their submissions. Valentine (1965) paid "tribute to Burt's generosity in helping others engaged in psychological research.... he has put many a young psychologist on his first path to advancement ... [his] elaborate discussions, which were most useful, but embarrassed me with the thought of the time he must have spent" (p. 20).

Joynson was most critical of Hearnshaw's claim about Burt altering the history of factor analysis to claim priority over some other contributors. Joynson noted that Burt differed from Spearman on some major issues; and could not ever be accused of disagreeing with Spearman only after the latter died. Burt, like many others, was not as convinced that the existence of a general factor excluded group factors, and Burt was far more interested in specialized abilities than Spearman. Joynson can find no reference to Pearson (1901) anywhere in the papers where Burt supposedly usurped the origins from Spearman. He argued that Hearnshaw failed to appreciate the enmity between Spearman and Pearson which Burt probably did not wish to enter, and that Burt was more influenced by Macdonell (1901) than Pearson. He contended that the reasons for Burt's renewed claims about the origins of factor analysis were probably related more to the publication of Thurstone's seminal work (1947) rather than to Spearman's death in 1945. Thurstone's book tended, in Burt's view, to ignore the contributions of Burt and other British factor analysts. This is indeed the case (compare Cronbach 1979) with respect to the importance of group factors as well as a general factor, and the priority of the invention of the centroid method. I do agree with Thurstone, however, that Burt's main omission (and possibly his lack of later influence) related to his lack of desire or belief in rotation as an aid to the interpretation of factors. Joynson clearly demonstrated that these claims of priorities can be traced to Burt's published work as early as 1909. I also note the lengthy footnote 3 in Burt (1949) which explained why so many researchers are "naturally unfamiliar with these early developments," why Pearson should get credit, and reference to a debate in 1910 between these various academics about priority.

There are still some who would contend that Pearson made the initial major contribution to the development of factor analysis, and there are peculiar histories of factor analysis that make Burt's beliefs seem mild (e.g., Schonemann, 1987; Steiger & Schonemann, 1976). Thurstone (1947) made sweeping claims about the British school (and generally omitted their influence), and Guilford (1936) omitted Burt altogether. Much of the historical debate can be traced to the differences between factor analysis (initiated by Spearman) and derivatives such as centroid and principal component analysis (initiated by Pearson). Controversy between whether there are, and the nature of, differences between these procedures still rages (Velicher & Jackson, 1990). Further, it needs to be noted that Burt was "foremost a practical applied psychologist," as Hearnshaw acknowledged (pp. 35, 92) and in this light many of Burt's claims are not altogether incorrect. Burt was the first, or among the first, to *use* some factor methods and this first *use* of factor methods is the consistent theme in his writings (e.g., Burt, 1968, p. 67).

Joynson concluded that there are so far no grounds for accusing Burt of historical falsification or lying; the critics have not established their case conclusively in a single instance, and "any delinquencies of which Burt might be at all plausibly accused appear to have been of a minor and relatively common variety, and come nowhere near the sensational frauds which have been so widely alleged" (p. 214). Hearnshaw claimed that the motive for the usurpation was self-aggrandizement, whereas Joynson suggested that it was more likely ambition or self-assertion, and when Burt's motives are labeled self-aggrandizement, then the accusers are already halfway to accepting his guilt. Burt, claimed Joynson, ascribed first place to another, was the most dependable supporter of Spearman's central idea (I do not agree with this), was pushing for British priority as well as his own, and could have pushed his claims more forcefully rather than in his modest and polite ways. Joynson questions why this belittling took 40 years to note and answers that it was because "there was no such campaign. Burt's views were put forward quietly, and unsensationally.... The notion of a mounting campaign against Spearman is a fiction of Hearnshaw's imagination. Burt was a sane man pursuing a rational purpose" (p. 125). This is not to contend that Burt was unbiased, as history is open to numerous shades of interpretation. The latter account of his own earlier papers appears to be substantially accurate, as he was never entirely derivative from Spearman. Hearnshaw has "painted the contrast too starkly, greatly exaggerating both the extent to which Burt was originally dependent on Spearman, and the nature and strength of his more independent behavior in maturity" (p. 109).

The Merits of the Cases

Joynson is convincing when he claims that the majority of Hearnshaw's arguments employ negative evidence: If Burt had possessed the data, he would have replied, but he did not reply so he probably did not possess them; moreover, one would have expected there to be traces of assistants, records of data, records in diaries, but such evidence is not found. Joynson's reaction to Hearnshaw's claims as to causes of Burt's indiscretions are convincing. It is difficult to support the notion that because Burt had Meniere's disease his critical judgment became impaired; he may have been more frustrated and thus may have become more careful. Joynson finds it remarkable, as do I, that Hearnshaw could claim that "there was in all probability an innate instability in Burt's psychosomatic make-up" (p. 277). A further claim of Hearnshaw borders on the ludicrous: "by reason of its exclusiveness (classical education) tends to induce feelings of superiority and elitism in those who are classically educated; and in a subtle way it is a training in pretence" because in preparation of producing pieces of Greek and Latin prose the perfect copy or fake achieves the highest marks. "For more than ten years of his early life Burt was subjected to this regular discipline" (pp. 263-264). While some of these claims by Hearnshaw may account for part of Burt's personality, it is beyond the scope of even a psychobiography to believe that this could explain why he allegedly cheated and so many others did not. Joynson's main contention was that Hearnshaw "makes all those assumptions which are consistent with the interpretation he wants to make, conveniently ignoring any pos-

sibility which might conflict" (p. 221). Hearnshaw selected only setbacks and does not account for the successes (such as honorary degrees, knighthoods, success of articles received, laudatory comments by colleagues).

Joynson is not believable in inventing excuses for Burt's failure to reply to requests for data (p. 175), claiming that the correlations for the measures of intelligence of the twins were not markedly stable when the mean difference appears to be very stable (mean difference = .043, sd = .038), that the misleading claims in a newspaper article may have been a consequence of Burt having misheard the question, that his critics are part of a left-wing, even Communist, conspiracy or part of a Maudsley cabal, and that the subject is moot because one cannot convict Burt of fabrication on suspicion, especially since he cannot defend himself.

Joynson is on the weakest ground when he tries to psychoanalyze Hearnshaw. Joynson is at pains to understand how Hearnshaw so misread Burt's character. Hearnshaw had already been working on the biography for some five years, and had written several chapters. Then along came the critics and Hearnshaw had to revise many opinions. This, claimed Joynson, added considerable pressure on Hearnshaw to complete his inquiries as soon as possible. Given the pressure, he hit upon the idea of striking a compromise between the anti- and pro-Burters, which "inevitably meant a verdict of guilty—not as guilty as the anti-Burters alleged, but not as innocent as the pro-Burters maintained" (p. 322). Joynson patronizingly commented that "a more leisurely timetable would have been wise" (p. 306). Joynson's attack on Hearnshaw is the worst feature of an otherwise excellent rebuttal.

The Reaction

There have been reviews of Joynson's book and the overriding opinion is bouquets for the detail and the way Joynson has corrected the little matters, accolades for having the bravery and time to try, and muted applause for edging the opinion slightly more to the side of innocence. The key players, however, are not convinced that Joynson has moved the opinion sufficiently to cross the line of innocence, and some have claimed that he is not even close.

Clarke and Clarke (1990a) are not convinced. They are not happy with Joynson's methods as he "unfortunately uses every available means, including gossip and innuendo, to cast doubt upon the integrity and scholarship of Burt's critics" and "in his zeal to vindicate Burt, Joynson has been less than scrupulous in his search for, and assessment of, the considerable documentary evidence," and he has "made many errors of fact" (p. 123). As in their earlier accusations, they provide no evidence for any of their claims. They now acknowledge that the changes supposedly made by Burt to their own work were "subtle" which greatly diminishes their case (Clarke & Clarke 1990b). Their reviews can be dismissed as shallow attempts to cast doubt on the integrity and scholarship of Joynson.

Beloff (1990), the organizer of the British Council symposium on Burt that aimed to "decently face the evidence before it," also is not convinced and contends that "it is a pity that [Joynson] does not take a fresh look at the problem, considered from a different social and ideological perspective." It is

difficult to imagine two more unlikely perspectives than Hearnshaw's and Joynson's. Beloff casts aspersions on Joynson's scholarship with condescending comments such as "it must be rare indeed for a contemporary work [i.e., Hearnshaw's] to receive such detailed textual analysis," "Joynson's emotive argument, induces despair more often than clarity," and she concluded that Burt "could not, nor can now, obscure the doubt, beyond reasonable doubt, that the later evidence was not empirical" (she means that Burt invented data). Her review is a poor apology for the 1970-1980s establishment of British psychology.

Two key players have modified their views. Eysenck (1990) who had already changed his views from innocent to guilty, now concludes that the most acceptable verdict is not guilty. He regards the case of the "missing ladies" as comic opera stuff and, correctly, does not take it seriously; he agrees with Joynson on the legitimacy of Burt's claims to priorities in factor analysis; but he believes that Burt did invent at least some of his twin data. Hudson (1989), who was responsible for the decision to burn Burt's archives and was an early accuser of Burt, was most concerned that Burt's transgressions were patently obvious but not discussed at the time, and he now seems to soften his view as to Burt's guilt.

Zenderland (1990) appears persuaded by Joynson's claims but concludes that Burt hardly leaves the courtroom with his reputation intact. She agrees with Joynson that Burt's methods may have been less than admirable, but they were short of criminal. She is not convinced by some of Joynson's claims about the kinship correlations, and believes that Joynson's "innocent" Burt emerges as an even less likable character than Hearnshaw's "guilty" Burt. Fortunately, this is not a popularity contest. Blinkhorn (1989) agreed with the claims about Burt's priorities regarding factor analysis and his elevation of Pearson, but was not convinced about Joynson's claims about the twin data. He found merit in the argument that it is harsh to judge pioneers by standards they ultimately set after their fieldwork was completed, and wonders if others are going to leave self-documented archives for future defenders. Fancher (1989) does not agree with the latter and sees less honor in Burt's deviousness. Some reviewers merely summarized the book (e.g., Ammons, 1990; Stein, 1989), approved the reinvestigation of the matter (Thieman, 1990; Tucker, 1990), or agreed that the question marks that Joynson has created deserve some response (Audley & Rawles, 1990; Plomin, 1989; Tysoe, 1990).

Hearnshaw (1990) and Joynson (1990) have also clashed since the publication of the latter's book. Hearnshaw accused Joynson of introducing no new evidence, not examining available evidence, not consulting him, not knowing Burt personally, and Joynson's book is of tedious length (Joynson is 27 pages shorter). These claims are either incorrect or not damning. Hearnshaw concluded that "Joynson's attempt to salvage Burt's reputation may be regarded as a total and dismal failure, and it fails above all because Joynson has simply not done his homework" (p. 64). Joynson (1990) reacted to these claims, but the details add little to the overall story. The editors of the British Psychological Society official magazine ruled in October 1990 that because they had already devoted considerable space to the Burt affair they

would not publish Joynson's rejoinder to Clarke and Clarke (1990b). They continued, however, to publish other letters on this subject in subsequent issues. They have also devoted much space to discussing Joynson's claim that Burt was the only psychologist to be knighted (the number now stands at seven plus a dame).

The Verdict

Both books are necessary, compelling reading. Both are extremely well organized and a sense of the author and subject are vitally present. Hearnshaw's biography has more flesh and bones about Burt and there is an overall persuasive argument as to Burt's guilt of fraud. At times Joynson protests too much and his attacks on the motives of the accusers detract from his overall message. Joynson has introduced the issue of whether Burt was guilty "beyond reasonable doubt," but he does not, nor do any reviewers, raise the issue as to whether only a taint or a full proof is needed for fellow researchers to discard Burt's data. Neither book is convincing such that a closed case exists. Like most historical mysteries, this is maybe how it should be.

So whom are we to believe? Joynson claimed that many believed Hearnshaw, because he enjoyed a reputation as a reliable and balanced historian. Further, he had reached a definite conclusion that satisfied most with the verdict of guilty as accused, as this "offered the best hope of a quick settlement and a speedy oblivion" but did not find Burt as guilty as some had been commenting or as innocent as others believed. The reviewers of Hearnshaw's book were laudatory, agreed with his conclusions (many were written by Burt's initial accusers, which Joynson claimed was reasonable except "in so far as their ill-concealed pleasure at the outcome outweighed their capacity for critical appraisal," p. 314), were brief, did not show much critical insight, or did not make much penetrating analysis of the book's strengths and weaknesses. Other reasons for support of Hearnshaw's biography included: Burt's most influential supporters changed their minds when the biography appeared; further papers, published after Hearnshaw, claimed more evidence against Burt; and the Council of the British Psychological Society condemned Burt on the basis of the biography. That the Council rushed into assuming guilt, particularly when two of the main accusers (Tizard and Clarke) were major office-bearers, still seems rather unseemly.

The impression of Burt often left by Hearnshaw and Joynson (and certainly reflected in the reviews and second-hand accounts) is that of slipshod practices, lack of detail, and a mind closed to and oblivious of opposite arguments. In preparation for this essay I have read most of Burt's work and I was struck by the detail, the lucidity, and the competence that emanates from it. We are dissecting the works of one of the clearest writers in psychology, and certainly one of the most literate and artful. Although Burt may have had an egomaniac demeanor, he is not alone in academia in this quality. Although I never met Burt, it seems reasonable to infer that his commanding manner was not endearing and his attacks not welcomed. During the past year I have read biographies of McLuhan, Blunt, Freud, and Mead and note many similarities in personality, and could name many living exemplars with

similar egomaniacal demeanors to Burt. Moreover, what would we make of an eminent psychometrician who promises to call his autobiography "Mean-square within"? Burt may not have been a sociable or clubable man (Hearnshaw, p. 183), but surely these indicators are not symptomatic of a criminal mind. It seems to be a serious charge to be accused of not being part of the club in Britain.

So was Burt guilty of anything? Burt could have collected some of his postwar data with the help of ex-students; he was honest about the low reliability of his data, and his caveats were readily available to those not convinced or willing to read his references; he probably received assistance from Conway and Howard although he used their names as by-lines, probably without their directly assisting him at the time of writing; he almost certainly changed contributors' submissions to the journal and bullied some ex-students and colleagues; he does deserve more credit for his pioneering use and developments in factor analysis; and he presented an argued (although arguable) case for the priority of Pearson's works. The case of the robust correlations is more difficult to reconcile, although it appears that it was a case of carelessness more than deceit or intention to commit fraud. To prove fraud, there must be a case for intention to deceive, and this is Hearnshaw's weakest point.

It does appear that Burt committed awkward, unskilled acts, but there is little evidence that he committed intentional deceit or fraud. Burt disappointed many, particularly with his devious replies, and at a minimum he can be accused of occasional slipshod work and that he took inappropriate liberties; at worst, we can ignore his work. To go further awaits more evidence.

The Moral

A number of lessons can be learned from this unfortunate affair. The need to be pure is the essence and there is no excuse for inventing data, changing conclusions (in any direction), or plagiarizing. It is not good enough to comment that comparatively few "can resist the temptation to indulge in a little faking, fishing, and even bare-faced prevarication, when things go wrong. But are not all of us tempted at times in much the same way? ... even scientists themselves often display different standards of truthfulness in different situations" (Burt, 1966b, p. 367). Students need many reminders of the importance of not comitting these sins. Starting as undergraduates they need to be reminded about the evils of copying from texts without reference, passing off another's argument as their own, not acknowledging sources (even if the source be their peers), trimming data to make it look more convincing, embellishing results, omitting aberrant values, misreporting methods, selecting only the "best" results, or hiding failed experiments. As I have traveled around universities during my present sabbatical leave, I have asked to see the guidelines and penalties for these offenses. The most notable feature is the lack of guidelines, their inaccessibility *prior* to committing any of these acts, and the minimal nature of the penalties. I am also aware of the difficulties of proving many of these issues to a panel of independent academics (and usually to fellow student juries) and thus see a stronger import for the clear—and available—delineation of the rules and procedures

for investigating fraud. It should also not be underestimated how difficult it is to prove fraud or innocence of it and/or to convince university authorities to investigate (Broad & Wade, 1982, document numerous cases where authorities were reluctant to investigate).

There is no doubt that Joynson is correct in asking for procedures to deal with false accusations. When one is wrongfully accused of fraud (or racism or sexism), it is extremely difficult to mount evidence against the claims, as most efforts lead to supporting the very claims. Most do not generate positive evidence and the onus is on the accused to prove innocence. In future, it is hoped that allegations are published along with the evidence and such claims open to serious academic debate, preferably prior to being popularized in newspapers. Any contentious area needs protection so that the advocates may say their piece, and there is a need for outlets for critical appraisal of these views. More published interchanges (such as exist in *Brain and Behavior* and *Interchange*) are needed.

More attention needs to be placed on popularizers of claims of fraud. The need for simplicity and everyday language can lead to misrepresentation, although it is not reasonable, as does Joynson, to criticize newspaper reporters for wanting to get a "big story." If the story is wrong, then appropriate skepticism can be generated. Editors need to get such stories correct, as the consequences of legal action can be a major threat. Recall that it was the same newspaper that broke the Burt story that bought the Hitler diaries: "It was the day the thunder of *The Times* turned into a whimper and *The Sunday Times* was forced to sniff the stench of self-deceit" (see Kohn, 1986, p. 159). I note a more recent "scandal" breaking in a newspaper concerning another academic, Bruno Bettelheim (Bernstein, 1990). It is hoped that the accusers learn much from the case for and against Sir Cyril Burt before these accusations are repeated as if true. What should be discussed is the procedures for investigating these claims and the ease with which these claims can be made on the basis of such little data.

Regardless of the pros or cons of the case against Burt, there are many instances of ludicrous claims in secondary sources. Broad and Wade (1982) commented that Burt's work was "riddled with glaring statistical errors, later shown to be a sign of wide-scale fraud ... even more remarkably, the critics of Burt's position also missed the red flags that leaped out from his disputed writings" (p. 203). They contended that Burt was a cheat, who "used his mastery of statistics and gift of lucid exposition to bamboozle alike his bitterest detractors and those who acclaimed his greatness as a psychologist" (p. 204). They portrayed Burt as lonely and embattled, with a tormented imagination.

One often suggested way to avoid fraud is to make raw data available, although it is recognized how difficult this is. Most universities I have been in inevitably discussed the beauties of archiving, and may have written the appropriate computer code to efficiently and effectively archive staff and students' data. After some time it becomes evident that the few who use these data do not justify the large cost of entering it and maintaining it.

There are other ways of screening data. Psychology, like most academic disciplines, has a long-standing basis on the peer review system. Although it

is fallible (Fiske & Fogg, 1990; Mahoney, 1977; Peters & Ceci, 1980), it needs to be strengthened rather than replaced. Reviewers and editors need to be more accountable for errors (such as failure to recognize plagiarism), more attention needs to be placed on comments relating to dishonesty (provided the comments are reviewed with particular attention to evidence), and more attention and credit need to be given to replications (e.g., who will start *The Journal of Replications*). Perhaps psychology needs more critics such as exist in the arts, music, and literature and more journals such as *Contemporary Psychology*. The most recent effort to begin such a journal in education failed (*Contemporary Education Review*). If psychology and education had more formal mechanisms for excellence (e.g., Nobel awards) then more rigor might be placed on probity, although such rewards might encourage undesirable behaviors. The move to assessing academics on the basis of performance indicators can also have this effect, as many nonproductive academics may be forced to publish even if they lack original or worthwhile ideas of their own.

There is also a need for more biographies in psychology. Boring's biographies are must reading for all aspiring psychologists, and his beginnings need to be continued and placed in front of all. There needs to be caution, however, in psychobiography, which is more likely to use inductive post-hoc justifications for the behavior. The motives for fraud are many, and there is much difficulty in accepting either Hearnshaw's explanations for Burt's behavior or Joynson's explanations for Hearnshaw's. To claim that cheating "may be like cirrhosis of the integrity" or the result of a pathological personality akin to that of compulsive gamblers (Luria, 1975) ignores simpler explanations of greed, fame, or the need for consistency. Merely depending on the memories of many, still living colleagues about a former teacher, mentor, or leader are not the most reliable indicator unless these views are placed in perspective. We need to recall the Greek historian Thucydides' maxim: Men are more anxious to be called clever than honest and suspect honesty of simplicity.

Afterword

In the long run, Cyril Burt's evidence on heredity does not matter. Others have replicated his work, typically with the same results (Bouchard, 1987, Bouchard & McGue, 1981; Osborne, 1990; Rimland & Munsinger, 1977; see also Snyderman & Rothman, 1988). His contributions to education have been made and generally are now surpassed, his statistical work ignored or replaced by alternative procedures, and as Cronbach (1979) noted "he was hopelessly out of touch with mainstream psychology" (p. 1392). So today it is possible to forget him while not changing any cherished views. He appears, unfairly, to be relegated to the waxworks alongside other infamous characters who have been vilified, tried, but have no opportunity to reply. To question all his work is ludicrous (is the mathematics in *Factors of the Mind* a figment of his fraudulent efforts?). It is of little interest to determine exactly when Burt's work becomes suspect (Hearnshaw cites 1943; Clarke & Clarke, 1980a, 1980b and Kamin, 1974 say all of it), although there is probably merit in the case that Burt would have made a good career decision

by dying at age 60. The 13th stroke of the clock brings into question the previous 12, and most researchers are not using any of Burt's work to support their claims.

Yet at the time Burt was influential, if not preeminently influential in Britain and the colonies. For right or wrong, Burt deserves to be considered among the most influential psychologists of this century. His deviations, if substantiated, are serious, but minor considering his direct influence on education and psychology. He made major contributions to the backward, gifted, mental testing, use of statistics in educational psychology, child guidance, delinquency, military psychology, subsequent criticism of the 11+, and as Hearnshaw noted "Burt's main achievements were in the field of individual, or differential, psychology; in some branches of mathematical psychology; and in theoretical psychology of a humanistic type—all areas about whose status psychologists have as yet reached no consensus" (p. 319). He deserves to have his portrait hung alongside the other pioneers—Spearman, Pearson, Thomson, and Garnett. There is no doubt, however, that his alleged deviations will be his lasting marks on psychology. He is not Hearnshaw's "mad professor" or Joynson's "happy warrior," he is now the marginal psychologist.

The message for present-day psychologists is clear: exactness, scrupulousness, and attention to detail are needed. "Be honest; never manipulate data; be precise; be fair with regard to priority; be without bias with regard to data and ideas of your rival; do not make compromises in trying to solve a problem" (Mohr, 1979, p. 45). Scandal is a honeypot for many. It is astounding the number of academics who have passed judgment on Burt, who sneer at the views he espoused, but who have never read Hearnshaw, or now, Joynson. These two books are required reading to be informed about the debate; a careful reading of some of Burt's work is necessary; and both need to be tempered with the realization that Burt has had his influence on education and now has an influence on biographical and ethical standards in psychology.

Note

1. I thank J. Rogers and H. Swaminathan for their comments on this review.

References

Ammons, C. (1990). Joynson, R.B. The Burt affair. *Psychological Reports, 55,* 351.

Audley, R.J., & Rawles, R.E. (1990). On a defence of Professor Sir Cyril Burt. *The Psychologist, 3,* 360-361.

Beloff, H. (1990). Review of Joynson: The Burt affair. *British Journal of Psychology, 81,* 395-397.

Bernstein, R. (1990, November 4). Accusations of abuse haunt the legacy of Dr. Bruno Bettelheim. *New York Times,* p. E6.

Blinkhorn, S. (1989, August 10). Was Burt stitched up? *Nature, 340,* 439-440.

Bouchard, T.J. (1987). The hereditarian research program: Triumphs and tribulations. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy.* New York: Falmer.

Bouchard, T.J., & McGue, M. (1981). Familial studies of intelligence: A review. *Science, 212,* 1055-1059.

Broad, W.J., & Wade, N. (1982). *Betrayers of the truth.* New York: Simon & Schuster.

Burt, C.L. (1909). Experimental tests of general intelligence. *British Journal of Psychology, 3,* 94-177.

Burt, C.L. (1949). Alternative methods of factor analysis and their relations to Pearson's method of "principal axes." *British Journal of Psychology (statistical section), 2*, 98-121.
Burt, C.L. (1954). The differentiation of intellectual ability. *British Journal of Educational Psychology, 24*, 76-90.
Burt, C.L. (1955). The evidence for the concept of intelligence. *British Journal of Educational Psychology, 25*, 158-177.
Burt, C.L. (1966a). The genetic determination of differences in intelligence: A study of monozygotic twins reared together and apart. *British Journal of Psychology, 57*, 137-153.
Burt, C.L. (1966b). Parapsychology and its implications. *International Journal of Neuropsychiatry, 2*, 363-377.
Burt, C.L. (1968). An illustration of factor analysis. In H. Butcher (Ed.), *Human intelligence: Its nature and assessment* (pp. 66-71). London: Methuen.
Burt, C.L. (1969). Intelligence and heredity: Some common misconceptions. *Irish Journal of Education, 3*, 75-94.
Clarke, A., & Clarke, A. (1980a). Comments on Professor Hearnshaw's balance sheet on Burt. In H. Beloff (Ed.), *Supplement to the Bulletin of the British Psychological Society, 33*, 17-19.
Clarke, A., & Clarke, A. (1980b). Review of Hearnshaw (1979). *British Journal of Psychology, 71*, 172-173.
Clarke, A., & Clarke, A. (1990a). Review of Joynson: The Burt affair. *British Journal of Educational Psychology, 60*, 122-124.
Clarke, A., & Clarke, A. (1990b). The Burt affair. *The Psychologist, 3*, 74.
Conway, J. (1958). The inheritance of intelligence and its social implications. *British Journal of Statistical Psychology, 11*, 171-190.
Cronbach, L.J. (1979). Hearnshaw on Burt. *Science, 206*, 1392-1394.
Eysenck, H.J. (1990, September 2). Sensitive intelligence issues. *Spectator, 263*, 26-27.
Fancher, R. (1989, November 9). Fixing it for heredity. *London Review of Books, 11*, 19-20.
Fiske, D.W., & Fogg, L. (1990). But the reviewers are making different criticisms of my paper! Diversity and uniqueness in reviewer comments. *American Psychologist, 45*, 591-598.
Glass, G.V., McGaw, B., & Smith, M.L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
Gould, S.J. (1981). *The mismeasure of man*. New York: Norton.
Grafton, A. (1990). *Forgers and critics: Creativity and duplicity in Western scholarship*. Princeton, NJ: Princeton University Press.
Guilford, J.P. (1936). *Psychometric methods*. New York: McGraw-Hill.
Hearnshaw, L.S. (1979). *Cyril Burt: Psychologist*. Ithaca, NY: Cornell University Press.
Hearnshaw, L.S. (1990). The Burt affair: A rejoinder. *The Psychologist, 2*, 61-64.
Hudson, L. (1989, November 3-9). Recalling a scapegoat. *Times Literary Supplement*, pp. 1201-1202.
Jensen, A.R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review, 39*, 1-123.
Jensen, A.R. (1974). Kinship correlations reported by Sir Cyril Burt. *Behavior Genetics, 4*, 1-28.
Joynson, R.B. (1989). *The Burt affair*. London: Routledge.
Joynson, R.B. (1990). The Burt affair: A reply. *The Psychologist, 2*, 65-68.
Kamin, L. (1974). *The science and politics of IQ*. Potomoc, NY: Erlbaum.
Kohn, A. (1986). *False prophets*. Oxford, UK: Blackwell.
Luria, S.A. (1975). What makes a scientist cheat? *Prism*, 16-18.
Macdonell, W.R. (1901). On criminal anthropometry and the identification of criminals. *Biometrika, 1*, 177-227.
Mahoney, M.J. (1977). Publication prejudice: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research, 1*, 161-175.
McAskie, M. (1978). Carelessness or fraud in Sir Cyril Burt's kinship data? A critique of Jensen's analysis. *American Psychologist, 33*, 496-508.
Mohr, M. (1979). The ethics of science. *International Science Reviews, 4*, 45.
Osborne, R.T. (1990). Cyril Burt's invariant kinship correlations: Another look. *The Mankind Quarterly, 31*, 163-170.
Pearson, K. (1901). On the systematic fitting of curves to observations and measurements. *Biometrika, 1*, 265-303.
Peters, D.P., & Ceci, S.J. (1980). A manuscript masquerade. *The Sciences, 35*, 16-19.

Plomin, R. (1989). Book review: The Burt affair. *Behavior Genetics, 19*, 739-740.
Rimland, B., & Munsinger, H. (1977). Burt's IQ data. *Science, 195*, 248.
Saint John-Roberts, I. (1976, November 26). Cheating in science. *New Scientist, 72*, pp. 465-469.
Schonemann, P.H. (1987). Jensen's g: Outmoded theories and unconquered frontiers. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy* (pp. 313-328). New York: Falmer.
Skanes, G.R. (1978). Correspondence. *Bulletin of the British Psychological Society, 31*, 201.
Snyderman, K., & Rothman, S. (1988). *The IQ controversy, the media and public policy*. New Brunswick, NJ: Transaction.
Steiger, J.H., & Schonemann, P. (1976). A history of factor indeterminacy. In S. Shye (Ed.), *Theory construction and data analysis in the social sciences* (pp. 136-178). San Francisco, CA: Jossey Bass.
Stein, D. (1989, August 24). The great IQ scandal. *Listener, 122*, p. 26.
Strachey, L. (1931). *Portraits in miniature*. London: Chatto & Windus.
Thieman, T.J. (1990). Burt. *Choice, 27*, 879.
Thurstone, L.L. (1947). *Multiple factor analysis*. Chicago, IL: University of Chicago Press.
Tucker, N. (1990, August 4). Mendacity or muddle? *Times Educational Supplement*, 19.
Tysoe, M. (1990, August 11). Burt the bogeyman. *New Statesman and Society, 2*, 30.
Valentine, C. (1965). Cyril Burt: A biographical sketch and appreciation. In C. Banks & P.L. Broadhurst (Eds.), *Stephanos: Studies in psychology present to Cyril Burt* (pp. 11-20). London: University of London.
Velicher, W.F., & Jackson, D.N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research, 25*, 1-28.
Zenderland, L. (1990). Burt again. *Science, 248*, 884-886.

The Alberta Journal of Educational Research *Vol. XXXVII, No. 3, September 1991, 277-294*

PERSPECTIVES

Kohlberg's Legacy—More than Ideas
An Essay Review

WILLIAM J. HAGUE
University of Alberta

Kohlberg's contribution to psychology was not only scholarly but personal, according to the contributors to this most intriguing tribute to an eminent psychologist and educator. As an introduction to Kohlberg's ideas on moral development this book is clear, practical and helpful. However in addressing the topic of the helping professions, some intepretations are unnecessarily restricted, while others, particularly in the case of implications for counseling, seem to let admiration lead to exaggeration. This essay concludes with an examination of the convergence of Kohlberg's legacy with character and virtue psychology.

Kuhmerker, Lisa, (1991), *The Kohlberg Legacy for the Helping Professions*. Birmingham, AL: Religious Education Press, 234 pages, Paper Cover, $14.95 U.S.

Some people give their name to ideas. We switch readily from the adjectival "Freudian" to saying, "That's Freud"; we say, too, "This is Rogers." Lawrence Kohlberg has become a name that is an idea. When we talk about a particular stage of moral reasoning, it is convenient to say, "That's Kohlberg." This name has taken its place with those of Freud, Skinner, and Rogers as a turning point in the history of psychology. Agree with them or not, one cannot deny that each of these men sent psychology and education off in a major new direction. Kohlberg had much to do not only with reshaping psychological theory, but also with giving practical direction to education, especially moral education. He addressed a need of his time in the United States with a philosophy and an empirically based theory that purported to walk the narrow path of educating morally without indoctrination while maintaining a principled morality. Kohlberg seemed to be the answer to the problem of maintaining moral standards without recourse to religion or talk of moral virtues. Content thus yielded to process; and this fit well with the American political and educational landscape of the time. Not too many years ago Kohlberg's famous moral dilemmas and his quickly accepted stages of moral development seemed to offer salvation not only to moral and religious educators, but to teachers in general. Now, a book entitled *The Kohlberg Legacy for the Helping Professions* offers this same salvation not

---

Bill Hague is a professor in the Department of Educational Psychology where he coordinates and teaches the graduate counseling program at the master level. He also teaches graduate and undergraduate courses in the psychology of religious and moral development and education. He is author of the book *New Perspectives on Religious and Moral Development.*

only to teachers but to a wider audience of those who help. The authors maintain that Kohlberg's influence goes beyond education, and his "legacy" extends much more broadly than some had imagined. What legacy did Kohlberg leave to education and to the helping professions, and what legacy does this book convey? These are the major themes of this review.

A Labor of Love

First a word about the tone of Lisa Kuhmerker's book. It is a scholarly work in the tradition of academic objectivity. But the unsuspecting scholar must beware; this is something quite rare in academia; this book is also a labor of love, rich with subjectivity. It is lovingly written by Lisa Kuhmerker and others who were close associates and obvious admirers of Lawrence Kohlberg over many years. We can expect it to be delightfully biased. It is replete with vignettes and personal impressions of Kohlberg the man. The central focus in it is clearly Kohlberg himself and what *he* has to offer the helping professions. The name of the man, as we have seen, can stand for the theory; in this book it stands for the theory, but the personal aura of the man permeates the whole. It is not without significance that the cover of the book shows a photo of Kohlberg with the sun shining through his hair in a distinctly halo-like effect.

Works by and about Kohlberg are numerous. Many are philosophical, theoretical. Many are psychological, both theoretical and practical, aimed sometimes at furthering research in moral education; at other times, directed to schooling. The niche that Kuhmerker, who is editor of the *Moral Education Forum*, carves out for this book is both theoretical and practical for the educator, the counselor, and the clinician. This book, after presenting the theory and the research on which practice is to be built, attempts to expand our conception of the practical dimensions of Kohlberg, Kuhmerker notes (p. 2) that this expansion is possible because, while Kohlberg's original research was limited to young people, his longitudinal studies and the research of others revealed that adult reasoning too shows stages and can be fostered by the challenges of adult life. This, she says, has had "a profound influence on counseling, pastoral care, clinical practice and programs in corrective institutions."

Almost any introductory psychology text in the last 20 years that has touched on moral development has featured Kohlberg and has contained at least a table giving his stages of moral development (sometimes canonized as "*The* Stages of Moral Development"). Psychologists and educators have a love affair currently with stages of development. Researchers and teachers are happy to use stages as neat containers for categorizing people. There is, it seems, a tacit promise that if we know the stage we will know the person, how to teach the person, and what to expect of him or her. The Kohlberg bandwagon was jumped on by many who saw his contribution as just that—a stage theory of individual development one could grab quickly and use immediately. But there was on Kohlberg's bandwagon much more than "neat" stages for categorizing people. Beyond were Kohlberg's conviction and lifelong dedication to the belief that schools need a democratic structure not only to develop individual moral reasoning, but also to foster a just society at

large. The issue was bigger, as big as the man himself, and that is what Kuhmerker tries to capture in this book.

Kuhmerker's practical aim is clear from the beginning:

> Our primary goal is to present Kohlberg's work clearly, so that the practitioner can be helped in his or her professional activity. With a few exceptions we do not focus on Kohlberg's critics. The annotated list of suggested readings for chapters 2-4 guides the reader to some of these critical evaluations. The bibliography at the end of the volume encompasses every one of the references by Kohlberg and his associates. Selected references by other authors are included. A doctoral student or researcher will find in this volume all the information he or she needs to decide whether or not Kohlberg's work is relevant to a projected research topic. The educator or counselor can see clearly whether or not the work of Kohlberg and his associates has relevance for his or her particular role in the helping professions. (p. 4)

Thus she makes a fairly modest but helpful claim: not that this book will solve all your problems; only that it will give you enough information to make some decisions. Those who have plowed through Kohlberg's often tangled style, written it seems by someone in a hurry to get on to the next project, will appreciate this intention to condense and clarify. With the above statement Lisa Kuhmerker succinctly outlines the goals of this volume, modestly eschewing any claims to be publishing the encyclopedia of Kohlbergiana, but aiming instead at clarity and usefulness to students and practitioners. The way these practical goals are pursued and the success with which they are accomplished are the substance of this review.

Contents and Organization

The book is divided into two major sections, the first theoretical empirical and methodological, the second oriented toward the educational practitioner. Lisa Kuhmerker begins this work by coming from her strengths: "More than twenty years of involvement in the field of moral development and long-standing personal contact with Kohlberg himself and with colleagues in the Association for Moral Education" (p. 3).

Part I — Moral Development: Kohlberg's Theory, Research Findings, and Methodology

This volume on moral development theory, research, and education begins imaginatively and appropriately with Lawrence Kohlberg's own life story—*My Personal Search for Universal Morality.* It starts with his high school days, his problems with conventional morality and formal religion, and his attempts at resolution by reading Dostoevsky's *Brothers Karamazov.* It continues with his role as an ex-merchant mariner running the British blockade of Palestine. As a student he worked as an attendant in a mental hospital, "seeing from the bottom up both the needs of individuals and the problems of social justice in that arena" (p. 13). He studied under Bettleheim, Rogers, and Gewirtz, noting that, "None seemed adequate to deal with the patient's moral problems and the justice pattern of the institutions" (p. 13). An incident in the hospital in which he felt that a patient's just rights were violated provided the occasion to "refocus (his) energies."

Such personal accounts, far from being irrelevant to understanding theory and research, slice, I think, to the heart of the matter. If theories were the product of free-floating brains, we would not need to understand the people who conceive and research them. But, products of the whole person that they are, our understanding of theories is enriched deeply by knowing something of the human being from whom they came. Kuhmerker is astute in acknowledging this and acting on it for our benefit. The presence of the personal and subjective in her book is a richness few others give us.

From this personal framework for experiencing Kohlberg's legacy, Kuhmerker moves to "a systematic examination of his theory, methodology and research findings within the context of twentieth-century philosophy and psychology" (p. 4). Chapter 2 (written by Uwe Gielen) introduces Kohlberg's prototypical and now famous "Heinz dilemma"—a man faced with the choice of stealing a drug to save his wife's life. Gielen outlines seven basic key terms and basic assumptions of cognitive developmental theory applied to the moral domain. The chapter surveys the refining process Kohlberg and his associates went through and multiple revisions of assessment strategies that have resulted. This alerts readers to be critical of which version of the Moral Judgment Interview is being used in a particular piece of research. Gielen has done a service for the reader unfamiliar with Kohlberg's cognitive developmentalism. He begins (p. 18) with fundamentals, contrasting Kohlberg's ideas with the psychoanalytic approach of Freud and social learning theory as represented especially in the field of morals by Hartshorne and May (1928/1930). The role of this chapter seems to be that of a primer of cognitive developmentalism à la Kohlberg, including not only the standard table of Kohlberg's stages of moral development but a "pro and contra" table giving helpful examples of responses to the Heinz dilemma at nine stages and substages of moral development. Altogether Gielen's is a clear presentation of the basics, including a seven-term lexicon, a discussion of autonomy and heteronomy, and a summary of Kohlberg's Platonic philosophical foundations; all this in a lucid style invaluable to a Kohlberg novice.

Chapter 3 describes Kohlberg's major research—the 20-year longitudinal study of the moral reasoning of American males. This is followed by an overview of research studies under the headings: "cross-cultural," "gender differences," and "the relationship of moral judgment to action."

Chapter 4 endeavors to unravel a complex Gordian knot in Kohlberg's work—the scoring system he used to measure moral judgment. It is followed by a survey of two methods that flowed from Kohlberg's system: Gibbs's Social Reflection Method and Rest's Defining Issues Test, comparing the appropriateness of the three systems. Because learning to score Moral Judgment Interviews requires long, complex, and sometimes expensive training, this chapter, following the practical aims of this book, describes the rationale and the process so that prospective researchers may judge the appropriateness of the approach and the feasibility of using it.

This chapter is an opus in itself and probably contains more information than the typical student really needs. The casual reader, then, should not be discouraged if he or she feels information overload.

But, on the other hand, one should beware of oversimplifications. Notable among the statements in this detailed historical section is the following claim on *hermeneutic objectivism*:

> The coding guide successfully establishes a balance between the philosophical-psychological need for deep interpretation and the scientific need for objectivity and reliability. It is this balance that makes the coding guide such a striking example of objective hermeneutics. The tender minded approach of the humanities and the tough minded approach of the natural sciences are reconciled in Kohlberg's psychological work. (p. 65)

One would prefer a more modest statement.

At this point, the end of part one, Kuhmerker inserts a thoughtful little gift for the faithful reader: "Suggested readings for chapters 2-4." It is the kind of thing a helpful thesis supervisor would give a graduate student—an annotated bibliography filled not only with relevant references, but also with the kind of insider comments that come from an informed mentor who knows the ropes and can lead the student directly to relevant sources.

Part II—Kohlberg for the Educational Practitioner

Chapter 5, Kuhmerker says,

> serves as an introduction to chapters 6 to 8 on educational interventions with which Kohlberg was involved. It puts the significance and limitations of the use of hypothetical dilemma discussions into perspective. It alerts the reader to the fact that Kohlberg had an influence on the fostering of moral development through traditional curriculum ... and that much of the last ten years of Kohlberg's professional life were spent in active participation in the creation of "just community" alternative school programs. Chapter 5 explores why a school might wish to transform itself into a "just community" and what the complexities of implementing such a goal might be. (p. 6)

Chapter 6 explores the topic of dilemma discussions, giving "how to" information to help the teacher incorporate discussions into the curriculum. Chapter 7 returns to the biographical theme of Kohlberg's work in the schools of Massachusetts and New York, while chapter 8 explores the theme of the just community in the voice of one of its students in dialogue with Kohlberg. Chapter 9 pursues the topic of the just community still further by assessing the moral culture of the school as influenced by the "just community" experience.

Part III—Kohlberg for the Religious Education Practitioner

Chapter 10 looks at the major question of the influence of religious education on moral education. After indicating that Kohlberg's focus on structure rather than content may have opened the way for religious educators of various persuasions to find relevance in Kohlberg's work, the chapter focuses on the work of Fowler (1984) and Oser (1980) who have used Kohlberg's stages as a model for their own design of faith and religious development.

Part IV—Kohlberg for the Counseling and Clinical Practitioner

Chapter 11 moves on to issues of clinical diagnosis and intervention, beginning with a description of "competing counseling traditions" current at the

time Kohlberg began his work and making, as we shall see, some rather remarkable claims for Kohlberg's influence on counseling and psychotherapy.

Chapter 12 begins to bring the book to its conclusion, pointing out genealogical strains passing from Kohlberg through Selman, Kegan, and Noam.

Part V—The Kohlberg Legacy for the Helping Professions

Chapter 13 focuses the volume toward its conclusion, with James Rest providing a backward and forward look at Kohlberg's stage concept and cognitive construction of social morality. The work concludes in chapter 14 as it began—with something of a personal tribute to Kohlberg in the form of warm reminiscences shared at the award dinner of the Association for Moral Education when it honored Kohlberg in 1982. This, Kuhmerker (p. 7) says, is "the Kohlberg legacy to his friends ... If it had been possible, this is the way he would have wished to reach you—the educator, the counselor, the clinician, the practitioner who cares about the moral life."

Critique

The foregoing is largely an overview of what the book is about, how it is organized, its stated goals and general tone. Obviously it covers a range of topics too extensive to review each in depth. A reviewer must choose just some to look at and critique in detail. These should be chosen for their intrinsic importance and not merely for their interest to the reviewer or because he sees them as being somehow within his ken. Perhaps the following topics for detailed exploration are navigation points of this reviewer as he plots his course between Scylla and Charybdis.

It struck me while reading this book that the later Kohlberg is much more interesting than the earlier one. Perhaps this is partly due to the fact that what Kohlberg proposed in his initial years as revolutionary now is, in some quarters, old hat and part of the psychological establishment. But Kohlberg was a scholar always open to new ideas and ready at times to admit his mistakes, even if it took him a while to come around to it. Witness the article by Kohlberg and Candee, "The Relationship of Moral Judgment to Moral Action" in Kurtines and Gewirtz (1984, pp. 52-73) where Kohlberg picks up Frankena's distinction between "deontic" (principled) judgments and "aretaic" (responsible) judgments, and, although in 1984 he can't agree with Gilligan's proposition that there are two independent moralities, one of rules and justice and one of responsibility and care, he does go on to admit that high level moral judgments can be made using "intuition," "heart," and "conscience." These are remarkable words coming from one of the most cognitive of cognitivists. His willingness to move even further on this issue toward the end of his life is illustrated in his contribution to Modgil and Modgil's (1986) monumental work, *Lawrence Kohlberg, Consensus and Controversy*. Fascinating too, as we shall see later on, is Kohlberg's creation of an enigmatic Stage 7 when for much of his academic life he had claimed Stage 6 as the ultimate moral stage. This book, as we will see, misses some of the interesting and challenging later developments of Kohlberg by concentrating on an introductory approach and avoiding much of the criticism that lit

Kohlberg's intellectual fire. It is a pity that a book that unabashedly sings the praises of the man behind the theory neglects to show adequately his openness to change.

The Man Behind the Theory

Perhaps, then, the first major criticism of this book should concentrate on that which is at once both its strength and its weakness, on that which opens a door, but leaves it only partly ajar. By this I mean its innovative presentation of the man within the theory. As such it does the reader and psychology in general a real service by helping fill in the vacuum that so many academic publications leave: What kind of a person is this who proposes these great theories? Are they the result of sheer cerebral activity? Or do the ideas not flow from who the person is, what he has experienced, and, perhaps most of all, what he has suffered in his life? The anecdotes in this volume are mostly in the absent-minded, lovable, dedicated professor area. As such they add color and warmth and lead the reader to greater depths of appreciation of the man, his humor, his dedication, and his ideas. But more could be done with this. Years ago, it seems to me, something very different began to come from the pen of a man who had written his cognitivism in abstract and sometimes muddied prose. It came out in poetic, mystical words that spoke about a dimension beyond mere reason. Kohlberg (1970) in simple, photocopied class notes, entitled *The Ethical Life, the Contemplative Life and Ultimate Religion—Notes Toward Stage 7* presented a deep glimpse of himself that was more obscured in his formal publications. He begins (p. 2) by replying to his students' two objections to his moral stages: their rational intellectualistic quality and their closed quality. "While the stages are closed," he says,

> much of the appeal of the concept of development is the appeal to an open and individual process based on personal experience not conceptual thought. I have tried to convince you that a solution to moral problems must ultimately be a closed and permanent solution; it must be universal and it must be rational. If it is not, you will be back in the quagmire of relativism.

"But," he continues,

> I appreciate your demand for a concept of development which is more open, more vague and individualistic, and more dependent upon personal, non-rational experience. So in talking to you of my Stage 7, I will not be ... drawing upon the authority of scientific findings and moral logic ... I will only be talking to you about my tentative efforts and those of others to make sense of experience, once you are aware of Stage 6 morality and scientific psychology ... I hope to take away from stage 6 excessive demands placed on it, the demand that it provide the answer to the problems of the meaning of life. (pp. 2-3)

He goes on (p. 3) to speak about "the logic of contemplation and "mystical logic." He says, (pp. 3-4) "Ultimately we all know that the deepest feeling is love, and the ultimate reality is life, and our problem for today is how we can love a life which makes burdensome demands upon us, and yet does not even console us by telling us the world is moral or just." To address these questions of ultimate meaning the "logic of action" encompassed by his six stages of moral reasoning fails because "life itself is not a rational process" (p. 21). Our

birth, our life, and our death are processes beyond reason, so he postulates a metaphorical Stage 7, which is religious in its mystical experiential, meaning-making dimensions.

> The pains of life which religion hopes to console are the pains of death, of unrequited love and separation and of injustice, and it attempts to console them by postulating the reality of something ideal or infantile. These pains, religion claims, reflect the discrepancy between our actual existence and some concepts of ideal or infinite limits in terms of which our existence falls short.
> To be pained by death is to conceive and value immortality,
> to be pained by injustice is to conceive and value ideal justice,
> to be pained by lack of love is to conceive of ideal love.
> To be pained by the finite and non-ideal is to recognize the reality of the ideal and infinite as something more than a relative human construction.
> If the world's injustice were nothing more than one's hedonistic pains, it would not be crushing. If the concept of eternity were nothing more than a fiction, death would be no pain. It is, then, human ideals which cause the existential pains, and their power is, in a sense, a sign of their reality.
> (1970, p. 36)

Kohlberg wrote this passage as prose, but I have taken the liberty of breaking the central lines of parallel structure into poetic lines because this passage, unlike so much of Kohlberg's scientific writing, sings like poetry. These words are too potent and too beautifully put to be just ideas. These are, in heightened contrast to scientific terminology, words of poetry and personal perception; one can intuit deep pain, tracing the inner journey of a person who has himself searched for ultimate meanings. One can pick up from Kohlberg's Stage 7 themes that these transcendent dimensions are not just the result of thinking, but of experiencing, and painful experiencing at that.

Kuhmerker includes a one-page allusion to Stage 7 in her chapter on religious education. But it takes a rather matter-of-fact, cooler approach to the topic. The references she quotes (Kohlberg, 1984) are Kohlberg's more formal, scientific, and less personal presentations on the topic. But there is too much in Kohlberg's willingness (or rather need) to break with sheer rationality and go beyond to deeper personal, religious dimensions to treat this as a mere interesting offshoot of a grand, rational theory. I think it tells us much about the bigness of the man open to see other possibilities, the depths of the man to experience new understandings, and the sensitivity of the man to know the reality of pain. One wonders why Kuhmerker does not recount this important dimension. Surely it is an important part of the Kohlberg legacy.

My own sparse encounters with Kohlberg have perhaps something to offer here. I attended the moral education conference in Leicester, England at which Kohlberg made one of his first personal forays into the British Isles. One afternoon, following his perennial critic, R.S. Peters, he spoke to the philosophers and educators in attendance. Judging by the mumblings in the audience, he did not in this initial presentation make a very positive impression. But that evening he gave a special extra lecture on his Stage 7. This lecture was much more personal and flowed almost lyrically. I was reminded

of the photocopied lecture notes (quoted above). After the lecture, I stopped him in the dining hall, prefacing my question with the eternal student lead: "May I ask you a personal question...?" To this I got the reply, "Try me." My question said in effect, "It seems to me that your Stage 7 is not just an academic adjunct, an afterthought tacked onto the classic cognitive stages, but it comes from your own personal religious experience." His reply was prompt and straightforward, briefly alluding to his return to the religion of his youth. His brief account of his own spiritual quest, given simply and openly there in the refectory of Leicester University, confirmed my intuition that such a profound appreciation of the transcendent dimensions of the person had been born not in the theorizing of academia, but in the disintegration and rebirth of real life. We talked for a while about the beauty that usually surrounds these mystical experiences, until someone interrupted with more mundane matters. One has to admire such personal integrity, depth, and sensitivity.

Gender Differences

Context is important here. Kohlberg was much criticized in his lifetime, notably by one of his former students Carol Gilligan for having concentrated his research and his conclusions on male subjects, favoring a cognitive approach deliberately abstracted from individual relationships in the interests, it seemed, of moral objectivity. Gilligan (1982) asserted a need to listen to the voices of women who, she claimed, tended to solve moral problems more in terms of caring and responsibility. The result was an appreciation of broader, personal dimensions of moral thinking. But there was danger that two camps would be built in the moral desert—one an ethic of principle, the other an ethic of care; one stereotyped as "male," the other as "female." Despite Gilligan's assertion that the two were complementary, the way was opened for opposition and exclusivism. Thus gender differences became a popular and hot topic.

Gielen, in the course of his review of research on moral reasoning (chapter 3) approaches this prickly question of gender differences. Consistent with his lucid style, he contrasts Kohlberg and Gilligan in these terms:

> While Gilligan's theory focuses on a broadly defined "moral self" Kohlberg's deontic conception of moral reasoning is restricted to moral decision making in situations involving the clash of competing moral claims ... In Gilligan's theory there exists a distinct danger that processes of moral judgment are assimilated to, and confused with personality organization.

It is interesting to note that it is all right for Kohlberg to be restricted to a deontic conception, but there is a "danger" in Gilligan looking at moral judgment in holistic personality terms. Are the disciples more orthodox than their master?

But Gielen must be given credit for the openness of his appreciation of his mentor. In the conclusion of chapter 3 he says:

> Kohlberg's earlier tendency to equate the moral domain with justice reasoning may have induced tunnel vision in some of his followers, but it kept his early theoretical and methodological efforts sharply focused on a fundamental dimension of morality. In his later years, he expanded his vision to include

> concerns for benevolence, beneficence, caring and responsibility, although his drive toward a broader conception of the ethical self was cut short by his death. The expansion of his moral vision is most salient in his educational efforts ... Without justice there could be no true morality for him. Caring and benevolence were important considerations in his own personal life, and they remained important for his interpersonal theory of morality. But he insisted that while care can not be expanded to include everybody—a commitment to justice can. (p. 59)

There is a risk that the ethic of care will be identified merely as a "feminist" approach to making moral judgments—a conclusion that would be constricting to the very notion of care as something *all* human beings should be capable of. Perhaps the task is not, in the interests of orthodoxy, to maintain Kohlberg's earlier, more restricted views, but to appreciate the more holistic approach he came to appreciate more, and to which the ethic of care and responsibility invites us. Perhaps the call is not to keeping orthodox guard, but to advancing into new ways of achieving moral objectivity that include but go beyond cognitivism.

Gielen's contributions to this book are generally strong, clear, and positive. They sustain Kuhmerker's aim to be maximally helpful.

One could go on at this point with a critique of those parts of Kuhmerker's book that extol the merits of Kohlberg's legacy for education in general and moral education in particular. But sufficient ink and paper have already been spent on this over the years, and further surveys of the pros and cons may be redundant. Instead I would like to look at those sections of the book under review that forge more into the hinterland of Kohlberg's legacy to the helping professions: notably religious education and counseling. It is appropriate here, while returning to the context of the whole book and before leaving the theme of gender issues, to note that in this volume nursing is conspicuous by its absence from major consideration as a helping profession. Do the authors feel that Kohlberg has no noteworthy legacy to offer this profession?

Kohlberg and Religious Education

Chapter 10 of this volume focuses in on the influence of religious education on moral development, attempting to dissect and analyze that sensitive area where religion and moral education intersect. This chapter does much to clarify the issues and underline the problems of Kohlberg and religion. But as I read Kuhmerker's contribution here, I see her using the word *religion* only in its formal, institutional sense, ignoring the distinction between that kind of religion and religion as personal and communal experience of the numinous, the transcendent—the kind of religion James (1958) explored in his *Varieties of Religious Experience,* and Jung (1938) celebrated throughout his lifetime. With the limited notion Kuhmerker utilizes, much of what she has to say does follow logically. But without the fuller understanding of the richness of religion, much of what could be said about the relation of religion and morality is lost.

Kuhmerker comes close to using this distinction under the aegis of Lee when she quotes him as saying:

> Kohlberg operationally defines religion as religious affiliation. Thus his data do not show that morality is autonomous from religion but from religious affiliation (Lee, 1980, p. 33). Furthermore Lee points out that if Kohlberg states that activities engaged in by the self are developmental, and "religion is actively engaged in by the self" then religion is developmental and can be investigated directly or indirectly in all its dimensions. (Lee, 1980, p. 336) (p. 165)

But instead of picking up this crucial distinction Lee has laid out in the clearest of terms, Kuhmerker chooses to ignore it on experimental grounds: "One might well ask however how Kohlberg would have studied the influence of religious education, other than to have correlated it with the process of being religiously affiliated" (p. 165). She then slips back into her overly simple dichotomies of religious versus secular by saying: "To look for other causes brings us back once more to assumptions of the 'extrinsic zap of grace' or the secular Piagetian-Kohlbergian assertion that the developing self is the constructor of its own philosophy" (p. 165).

This failure to see religion as developmental, something in process, and to identify it instead almost solely with content leads Kuhmerker to shift words when she looks at religion as process and call it "faith development" in the light of Fowler's (1984) work. Even the word *faith* begs for clarification it does not get in this book; *faith* can range in meaning from the broad process concepts of Fowler (1984) and Tillich (1958) as a search for ultimate purpose and meaning in life to a most conservative view that "the faith" is a collection of specific tenets to be learned and held to, unchanging for a lifetime.

Overall, Kuhmerker, following the tenor of this volume, presents the potential religious educator with a rather concise summary of the interaction of Kohlberg's concepts of moral education and a variety of important names and approaches to religious education. For beginning religious educators this chapter nicely surveys the relevant Christian and Jewish religious educators. As a short cut to important names like Fowler, Oser, Lee, Groome, and Chazan, this chapter is a précis of introduction to some contemporary religious education thinking. It concludes with a typically Kuhmerkian device—six points of practical advice entitled: "Guidelines for Parents and Educators Who Want to Foster Religious Awareness and Commitment" (pp. 168-169). They are based on the premise that fostering moral reasoning "enables children to reason about all issues, including religious issues, in increasingly mature ways" (p. 168). One cannot quarrel with the importance of encouraging reasoning in moral education, particularly when some attempts at "religious" education are replete with examples of deliberately ignoring reason. But the theme here is Kohlberg's "reasoning is enough." One wonders if this is a real reflection of the later Kohlberg, given the broader view of moral choosing he began to take particularly in regard to religious issues (Kohlberg, 1974, 1984; Kohlberg & Candee, 1984).

Kuhmerker emphasizes (p. 157) that, "It is the public school's role to foster the capacity to reason about moral issues without 'teaching' the belief system of any organized group." This is true. But I think there is a danger of presuming that the obverse side of this coin is also true: that the religious school is always and everywhere interested only in teaching its own belief

system. There have been centuries of traditional catechising that has meant "conversion," but recent statements by some mainline churches show a promising emergence from a narrow position of rightness to a more ecumenical and humanistic concept of the whole person in process. Statements of ecclesiastical purpose are not, of course, necessarily followed by universal implementation, but they do point to the error of presuming that religious education must of necessity be indoctrination in narrow content. Reasoning can be and is encouraged in some religious settings with or without Kohlberg. Indeed, the religious setting is the most suitable situation for reasoning to flourish in the context of the whole person. Kohlberg himself (1974) was well aware of that.

Hyde (1990) has written an encyclopedic survey of empirical research in religion in childhood and adolescence, including an entire chapter on religion and morality in adolescence. It is interesting that in the over 1,800 references in the book, not one acknowledges Kohlberg's research. Where his name does appear, it is a passing reference to his foundational philosophy linked with Piaget or his influence on Oser's stages of religious development. Perhaps this says something about Hyde's British origins or about Kohlberg's North American popularity. Such an omission is, I think, regrettable, failing to recognize the monumental contribution Kohlberg with all his research flaws and sometimes questionable philosophy actually did make to education, moral education, and indeed to religious education.

Kohlberg's "Contribution" to Counseling

Richard Hayes authored chapter 11, "Counseling and Clinical Implications of Lawrence Kohlberg's Developmental Psychology." At this point in the book, admiration for the seeming universal importance of Kohlberg seems to take over from objective fact. From the introduction, the unwary reader might conclude that 1969 marked the dawning of a new age of counseling when that profession, thanks to the publication of a Kohlberg article (presumably the "Stage and Sequence" paper in Goslin, 1969, although that is not clear) suddenly became capable of a "full appreciation" (p. 173) of a developmental view.

One claim leads to another and we have Hayes (p. 173) generalizing that, "Prior to Kohlberg's seminal work in moral development counselors generally accepted one or the other of two competing views of human development." (Already it sounds too simple.) Hayes then explains that Kohlberg distinguished the "cultural transmission model" from the "romantic model." As someone who was a counselor educator in the 1960s and 1970s, I tried to locate myself in this dichotomy. Caught on the horns of this dilemma, I felt less comfortable on the behavioristic horn and so found myself labeled a romantic. Hayes then switches to his own terminology (without warning the reader who must fish in the words for his or her own continuity) to a bifurcation labeled "environmentalist" and "maturationist." Even though I profess a concern for the environment, I thought that I probably belonged in the maturationist category. And that seemed to fit until I came to the words: "For the maturationist, change is attributed to growth and differentiation. Environmental factors are recognized, but only as *placing limits* [emphasis

mine] on individual development, not as directing its course" (p. 174). There I felt I must part company with this limiting "maturationist" view of development and whatever it had to say about counseling.

To add to the confusion, Hayes then goes back to Kohlberg's original "cultural transmission" versus "romantic" terminology, presumably to prepare the reader for quotes from the master. These quotes from Kohlberg and Mayer (1972) represent the thinking of Kohlberg almost two decades ago; they do not reflect the sophistication of thinking and distinctions that existed in counseling theory 20 years ago, never mind today! For example, Hayes (p. 175) quotes statements from Kohlberg describing the counselor as "value-free" or "value-neutral." It is not clear from the quotations at this point whether Kohlberg is endorsing the idea or not; but later we will find that Kohlberg appropriately denounced value neutrality in counseling. Hayes is giving us a watered-down and confusing Kohlberg.

It is remarkable how Hayes gets into the value neutral problem in the first place. He does it by claiming that both the romantic and industrial psychology strategy of the cultural transmissionists come from a reductionism that betrays what Kohlberg has called a "bag of virtues" (p. 175). I cannot see how people who profess the teaching of specific virtues can have much to do with value neutrality except by way of contrast.

Another counseling issue Hayes touches on is the role of crisis in development. He notes Kohlberg's recognition of critical periods in development. But he then goes on to say in the next sentence (slipping into a narrower meaning of the word *crisis*) that, in the maturationist view of counseling, "Treatment is reduced to crisis intervention or the prevention of future crises." Again this may have been prevalent thinking 20 years ago, to which period Hayes is regrettably continually taking us back in order to show the implications Kohlberg for counseling today. But today we see a crisis as an opportunity, something to be grasped and used for growth, a disintegration of a lower level so that it can be replaced by higher level functioning (Dabrowski & Piechowski, 1977). Rather than these stale statements from the distant past, the theme that moral development comes about not *despite* but *through* moral crises (Hague, 1976) is much preferable, more appropriate, and a more faithful expression of the radical Kohlberg legacy.

Finally (p. 176), Hayes tells us that neither of these two categories, the behaviorist or the romantic, to which counselors of a quarter century ago paid their allegiance is the right way. Kohlberg has instead offered the "tradition of progressivism which encourages the nourishment of the individual's natural interaction with a developing society or environment, and a cognitive-developmental psychology to explain the nature of human development" (p. 176). In the section on basic assumptions, a naive reader may wonder at the genius of Kohlberg, who is purported to have constructed an elaborate theory of development that seems to incorporate the best of existentialism's meaning-making, behaviorism's social constructivism, symbolic interactionism, and organismic theories' attention to qualitative reorganization in the tradition of Werner and Kaplan. At the summit of all this, Kohlberg emerges, of course, with a cognitivist Piagetian approach, and does

so in a progressive tradition, following Dewey. It is a bit much for anyone to synthesize.

Instead of the elaborate hermeneutics Hayes uses to dig all these facets of Kohlberg primarily from the depths of a single paper, one would appreciate more an elaboration of a few key points in Kohlberg's psychology and what counselors can learn from them. For example, role taking is surely a major theme of Kohlberg's psychological and educational thinking. It is essential to development in the moral atmosphere of the school. The counselor has in his or her vocabulary the word *empathy* and he or she may spend a lifetime trying to get it right so that there is involvement with a client without getting all mixed up in the client's problem. If a researcher could find that Kohlberg had something to say about this empathy in the context of his experience with role-taking, he or she could indeed make a fine, *original* contribution to counseling.

Would it be too simple to say that Kohlberg was a moral philosopher and educator, and that consequently he may have something to tell us about the role of moral conflict in counseling? Clients frequently have moral problems mixed in with other personal concerns. Would it be simplistic to suggest that Kohlberg's legacy is to insist that, instead of skirting moral issues, one must confront them? Then to say that this confrontation, though painful, can be growthful? And then to say that Kohlberg gives us at least one practical way of handling personal issues—with reason?

Even more basically, one could take Kohlberg's fundamental cognitivism and find it akin to Ellis's Rational Emotive Therapy, but with a dimension that Ellis does not seem to have: an appreciation that thinking is not just rational or irrational, but that reasoning itself is multileveled. A client would, with this appreciation, be encouraged not just to think rationally, but to think at a higher level. The concept of moral levels would be a major contribution to Rational Emotive Therapy that has been wanting in this area.

After reading this chapter, one is not as convinced, as the author seems to be, of the great innovative insights into counseling offered by Kohlberg, especially when so much of it depends on a single article by Kohlberg and Mayer published in 1972. It shows, instead, that Kohlberg understood to some extent what was going on in counseling at the time, pigeonholed counseling in rather generic categories, and presented his own alternative approach as part of the progressive tradition of Dewey in the cognitive developmental framework of Piaget. Altogether, this chapter is a tribute to Kohlberg's broadness of view, his scholarship, and his personal concern for people and for counseling (Kohlberg, 1975; Kohlberg & Wasserman, 1980). But he is far from being the originator of any special school of counseling. The most that can be said is that his assumptions can support and supplement those of some others who have better worked out and more elegant theories of counseling.

A Question of Virtue

Kohlberg's early writings (1958, 1963, 1964) were highly critical of Hartshorne and May's (1928/1930) landmark research on character education. They had tried to assess the success of moral education in terms of

virtues practiced in real-life settings. Little was empirically supported. The shakiness of this research led Kohlberg, in his concern for an objective morality, to shun traditional character education, loaded as it was with content, and to look rather in the field of cognitive process. The term *bag of virtues* took on a scornful note. To right the imbalance of former years of character education (often in a religious setting) Kohlberg focused deliberately on the structure of reasoning about social justice rather than its content. This was historically opportune and suited the needs of public education in the United States. It fit well with progressivism and a democratic society eager to encourage individual self determination, offering at the same time the objectivity of principled behavior founded now not on many virtues, but on one—justice.

This process emphasis promised to avoid the indoctrination of content identified largely with private and religious schools. Interestingly, Kuhmerker (p. 157) even claims that this emphasis on process has recently come full circle to make Kohlberg's work "relevant to theologians and religious educators of widely divergent points of view." But behind this thinking is an assumption about religious education—that it is invariably obsessed with content, and must come full circle, back on its own tail so to speak, to appreciate the role of process in religious education. Such stereotypes are far from the reality of contemporary religious education which is aware of the importance of a happy combination of process and content. Content is an important part of moral education, and moral education is in a healthier condition when content is part of the open curriculum rather than buried in a hidden curriculum. In political situations demanding a split between church and state, some moral education approaches have become preoccupied with freeing themselves from content in an effort not to be "religious." But to say that moral education can be truly content-free is tantamount to the protestations of psychologists a few years ago that their counseling must be value-free. We have grown enough in the field of counseling to admit that values are present and active. It is time in moral education to admit that virtues are taught, consciously or unconsciously, and their practice encouraged as good habits of behavior. Character education has not for the last 20 years been the "in" terminology of moral education. Concern for an objective morality has concentrated on a Kantian categorical imperative and the role of reason in discovering a universal justice. Kohlberg has been the eminent champion of this cause. But recently other approaches are being explored. One can see it in the writings of Kupperman (1988) and Nussbaum (1988) who wish to develop the Aristotelian tradition of the possibility of moral objectivity in a world that recognizes a multiplicity of virtues. *Character education* and *virtue* are words coming over the horizon, not as the dark monsters they have in the past been taken for, but as a sunrise, proclaiming a new day in moral education (Hague, 1990).

Perhaps it seems strange to juxtapose *new* and *virtue*, for is virtue not an ancient word encrusted with medieval ideas of the indoctrination of God's will? Is it not contrary to all the autonomous rationality Kohlberg challenged? MacIntyre's *After Virtue* (1981) seems to lament the death of virtue. But authors such as Meilander (1984) and Barden (1990) have reminded us

that virtue is, after all, the moral strength of a good habit. If we can get away from seeing habit as a low-level, knee-jerk type response and perceive it instead as something consciously learned, developed, and strengthened with real-life practice, it may well serve again as the idiom for moral education. The work of Hauerwas (1985) has promised a new bridge between psychology and theological ethics. Important social commentators such as Lasch (1991) are drawing our attention to the need of specific virtues in our time, particularly the need for strong hope to replace shallow optimism.

Gielen (p. 49), as we have seen, warned us of the "danger" of "confusing" moral judgment with personality organization. Dykstra and Parks (1986) do not seem to share this fear:

> Neither Piaget nor Kohlberg intends to provide a theory of ego or personality development. Both, therefore, have approached the task of identifying the forms of reason or logic characteristic of different "stages" in human thought without making a critically important distinction: They have not attended to the differences between constitutive knowing in which *the identity or worth of the person is not directly at stake* and constitutive knowing in which it is. This has meant that Kohlberg has avoided developing a theory of the moral self, of character, or of conscience. Strictly speaking, his stages describe a succession of integrated structures of moral logic. He has given very little attention to the fact that we "build" ourselves through choices and moral (self-defining) commitments. His theory, for understandable theoretical and historical-practical reasons, has not explicated the dynamics of the inner dialogue in moral choice between actual and possible selves. (p. 22)

The task of the post-Kohlbergian moral educator will be to explore the problems and the possibilities of character education, how virtue might be taught and practiced, and how we might evaluate it with objectivity. This, interestingly, can be done in keeping with those concerns of the Kohlberg legacy which truly are most basic.

Conclusion

What, then, can be said of Kohlberg's legacy and Kuhmerker's efforts to pass it on? That Kohlberg left a rich legacy to pass on to psychology and moral education needs no further proof than this volume provides in the act of passing that legacy on. If one wishes to explore the fundamentals of Kohlberg's moral development from theory to application in the classroom and research, this book is probably the clearest, most comprehensive, and helpful publication in existence. Likewise, if one is interested in a glimpse at the man behind the theories, this book will provide that insight with a little help from his friends.

If one is interested in the implications of Kohlberg for religious education, this book is helpful and interesting. But, given the restricted parameters imposed on this field by its authors, one needs to be wary. However, if one is interested in exploring the implications of Kohlberg for the helping professions of counseling and psychotherapy, the sections of this volume devoted to that topic are hopelessly inadequate and out of date and naively distort the Kohlberg legacy out of all proportion. Kohlberg deserves a more balanced appreciation of how his ideas can be incorporated into the more sophisticated

theories of counseling that do exist and could benefit from his basic insights without pretense that this part of his very real legacy is more than it really is.

If one is only interested in Kohlberg himself, what kind of a man he was, this book is probably the best there is. But this must be taken with the proviso of a certain bias because it is written lovingly by friends and associates who hold his name in benediction.

References

Barden, G. (1990). *After principles.* Notre Dame, IN: University of Notre Dame Press.

Dabrowski, K., & Piechowski, M. (1977). *Theory of levels of emotional development* (2 vols.). Oceanside, NY: Dabor.

Dykstra, C., & Parks, S. (Eds.). (1986). *Faith development and Fowler.* Birmingham AL: Religious Education Press.

Fowler, J. (1984). *Becoming adult, becoming Christian.* San Francisco, CA: Harper & Row.

Gilligan, C. (1982). *In a different voice.* Cambridge MA: Harvard University Press.

Goslin, D. (Ed.). (1969). *Handbook of socialization theory and research.* Chicago, IL: Rand McNally.

Hague, W. (1976). Positive disintegration and moral education. *Journal of Moral Education, 5,* 231-240.

Hague, W. (1990). Shared values for shared life meaning: The transcendent dimension of the global village. In B. Holyst (Ed.), *Mental health in a changing world* (pp. 27-38). *Warsaw: Polish Society for Mental Health.*

Hartshorne, H., & May, M.A. (1928/1930). *Studies in the nature of character, Vol. 1: Studies in deceit; Vol. 2: Studies in service and self-control; Vol 3: Studies in organization of character.* New York: Macmillan.

Hauerwas, S. (1985). *Character and the Christian life.* San Antonio, TX: Trinity University Press.

Hyde, K. (1990). *Religion in childhood and adolescence.* Birmingham AL: Religious Education Press.

James, W. (1958). *Varieties of religious experience.* New York: New American Library.

Jung, C.G. (1938). *Psychology and religion.* New Haven, CT: Yale University Press.

Kohlberg, L. (1958). *The development of modes of thinking and choices in the years from 10 to 16.* Unpublished doctoral dissertation, University of Chicago.

Kohlberg, L. (1963). Moral development and identification. In H. Stevenson (Ed.), *Child psychology. 62nd yearbook of the National Society for the Study of Education* (pp. 277-332). Chicago, IL: University of Chicago Press.

Kohlberg, L. (1964). The development of moral character and ideology. In M.L. Hoffman & L.W. Hoffman (Eds.), *Review of child development research* (vol. 1, pp. 383-431). New York: Russel Sage Foundation.

Kohlberg, L. (1970). *The ethical life, the contemplative life and ultimate religion—Notes toward stage 7.* Photocopied notes for Social Sciences 154.

Kohlberg, L. (1974). Education, moral development and faith. *Moral Education Forum, 4,* 1.

Kohlberg, L. (1975). Counseling and counselor education: A developmental approach. *Counselor Education and Supervision, 14,* 250-256.

Kohlberg, L. (1984) *Essays on moral development, Vol. II: the psychology of moral development* (pp. 52-73). New York: Harper & Row.

Kohlberg, L., & Candee, D. (1984). The relation of moral judgment to moral action. In W. Kurtines & J. Gewirtz (Eds.), *Morality, moral behavior and moral development.* New York: Wiley.

Kohlberg, L., & Mayer, R. (1972). Development as the aim of education. *Harvard Educational Review, 42,* 449-96.

Kohlberg, L., & Wasserman, E. (1980). The cognitive-developmental approach and the practicing counselor: An opportunity for counselors to rethink their roles. *Personnel and Guidance Journal, 58,* 559-567.

Kupperman, J. (1988). Character and ethical theory. In P. French, T. Uehling, & H. Wettstein (Eds.), *Midwest studies in philosophy, Vol. XIII: Ethical theory: Character and virtue* (pp. 115-125). Notre Dame IN: University of Notre Dame Press.

Kurtines, W., & Gewirtz, J. (Eds.). (1984). *Morality, moral behavior and moral development.* New York: Wiley.

Lasch, C. (1991). *The true and only heaven.* New York: Norton.

Lee, J.M. (1980). Christian religious education and moral development. In B. Munsey (Ed.), *Moral development, moral education and Kohlberg.* Birmingham AL: Religious Education Press.

MacIntyre, A. (1981). *After virtue.* Notre Dame IN: University of Notre Dame Press.

Meilander, G.C. (1984). *The theory and practice of virtue.* Notre Dame, IN: University of Notre Dame Press.

Modgil, S., & Modgil, C. (1986). *Lawrence Kohlberg: Consensus and controversy.* Philadelphia, PA: Falmer.

Nussbaum, M. (1988). Non-relative virtues: An Aristotelian approach. In P. French, T. Uehling, & H. Wettstein (Eds.), *Midwest studies in philosophy, Vol. XIII: Ethical theory: Character and virtue* (pp. 32-53). Notre Dame IN: University of Notre Dame Press.

Oser, F. (1980). Stages of religious judgment. In C. Brussel (Ed.), *Toward moral and religious maturity.* Morristown NJ: Silver Burdett.

Tillich, P. (1958). *Dynamics of faith.* New York: Harper & Row.

For Contributors

Please prepare manuscripts in accordance with these guidelines.

Format

- Manuscript in triplicate.
- Typewritten or printed, double-spaced throughout.
- Abstract of approximately 100 words on a separate sheet.
- Manuscript not to exceed 6,500 words excluding graphics.
- Author's name and affiliation on a separate title page; title only on first page of manuscript to ensure anonymity in the review process.

Referencing

- Sources in parentheses after each reference (direct or otherwise), giving author's name (unless mentioned in text), year of publication, and page number(s).
- Direct quotes of 40 words or less in double quotation marks incorporated into text; those longer than 40 words in indented block format.
- All sources listed alphabetically at the end of the manuscript under the heading *References* using the APA style.
- Explanatory notes, numbered consecutively and marked in the text with superscript numerals, may appear before the References under the heading *Notes*; any citations in notes follow the same format as other references.

Graphics

- Each table and figure on a separate page at the end of the manuscript.
- Figures, charts, and diagrams in camera-ready form.
- Number tables in Arabic numerals and center over each heading, e.g., *Table 1.*
- Number figures in Arabic numerals before the caption at foot of each graph, e.g., *Figure 1: ...*
- Indicate placement of figures and tables in text, i.e., *Insert Table 3 here.*

Style

- *AJER*'s editorial style conforms closely to the *Publication Manual of the American Psychological Association.*
- Editorial changes may be made to manuscripts.
- For spelling, consult *Webster's New Collegiate Dictionary;* spelling in quoted material remains as in the original.

Diskettes

If manuscript is accepted for publication and has been prepared using a Macintosh, IBM, or IBM compatible microcomputer, please send diskette. Microsoft Word, MultiMate, WordPerfect, WordStar, Xerox Writer, XyWrite, and DisplayWrite 3 only are acceptable.
Important: Please indicate program used.

Submission to *AJER* implies the manuscript is not being considered for publication elsewhere.

AJER is indexed in the *Alberta Education Index, Canadian Education Index, Current Contents/Social and Behavioral Sciences, ERIC/Current Index to Journals in Education,* and *Social Science Citation Index*; appropriate articles are abstracted in *Cognitive Development Abstracts, Educational Administration Abstracts, Educational Technology Abstracts, Family Abstracts, Language Behavior Abstracts, Multicultural Education Abstracts, Psychological Abstracts, Research into Higher Education Abstracts, School Organization and Management Abstracts,* and *Sociology of Education Abstracts*, and reviewed by the *Journal for Research in Mathematics Education.*